Gaussian Processes

Neil D. Lawrence

GPRS
9th August 2013

Outline

Multivariate Gaussian Properties

Distributions over Functions

Covariance from Basis Functions

Basis Function Representations

Outline

Multivariate Gaussian Properties

Distributions over Functions

Covariance from Basis Functions

Basis Function Representations

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Multivariate Consequence

- If

$$\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

Multivariate Consequence

- If

$$\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

- And

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

Multivariate Consequence

- If

$$\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

- And

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

- Then

$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{W}\boldsymbol{\mu}, \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^\top\right)$$

Multivariate Regression Likelihood

- Noise corrupted data point

$$y_i = \mathbf{w}^\top \mathbf{x}_{i,:} + \epsilon_i$$

Multivariate Regression Likelihood

- Noise corrupted data point

$$y_i = \mathbf{w}^\top \mathbf{x}_{i,:} + \epsilon_i$$

- Multivariate regression likelihood:

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \mathbf{w}^\top \mathbf{x}_{i,:}\right)^2\right)$$

Multivariate Regression Likelihood

- Noise corrupted data point

$$y_i = \mathbf{w}^\top \mathbf{x}_{i,:} + \epsilon_i$$

- Multivariate regression likelihood:

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \mathbf{w}^\top \mathbf{x}_{i,:}\right)^2\right)$$

- Now use a multivariate Gaussian prior:

$$p(\mathbf{w}) = \frac{1}{(2\pi\alpha)^{\frac{p}{2}}} \exp\left(-\frac{1}{2\alpha} \mathbf{w}^\top \mathbf{w}\right)$$

Posterior Density

- Once again we want to know the posterior:

$$p(\mathbf{w}|\mathbf{y}, \mathbf{X}) \propto p(\mathbf{y}|\mathbf{X}, \mathbf{w})p(\mathbf{w})$$

- And we can compute by completing the square.

Posterior Density

- Once again we want to know the posterior:

$$p(\mathbf{w}|\mathbf{y}, \mathbf{X}) \propto p(\mathbf{y}|\mathbf{X}, \mathbf{w})p(\mathbf{w})$$

- And we can compute by completing the square.

$$\log p(\mathbf{w}|\mathbf{y}, \mathbf{X}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2 + \frac{1}{\sigma^2}\sum_{i=1}^{n} y_i \mathbf{x}_{i,:}^\top \mathbf{w} - \frac{1}{2\sigma^2}\sum_{i=1}^{n} \mathbf{w}^\top \mathbf{x}_{i,:}\mathbf{x}_{i,:}^\top \mathbf{w} - \frac{1}{2\alpha}\mathbf{w}^\top \mathbf{w} + \text{const.}$$

$$p(\mathbf{w}|\mathbf{y}, \mathbf{X}) = \mathcal{N}\left(\mathbf{w}|\boldsymbol{\mu}_w, \mathbf{C}_w\right)$$

$$\mathbf{C}_w = (\sigma^{-2}\mathbf{X}^\top \mathbf{X} + \alpha^{-1})^{-1} \text{ and } \boldsymbol{\mu}_w = \mathbf{C}_w \sigma^{-2}\mathbf{X}^\top \mathbf{y}$$

Bayesian vs Maximum Likelihood

- Note the similarity between posterior mean

$$\boldsymbol{\mu}_w = (\sigma^{-2}\mathbf{X}^\top\mathbf{X} + \alpha^{-1})^{-1}\sigma^{-2}\mathbf{X}^\top\mathbf{y}$$

- and Maximum likelihood solution

$$\hat{\mathbf{w}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$$

Marginal Likelihood is Computed as Normalizer

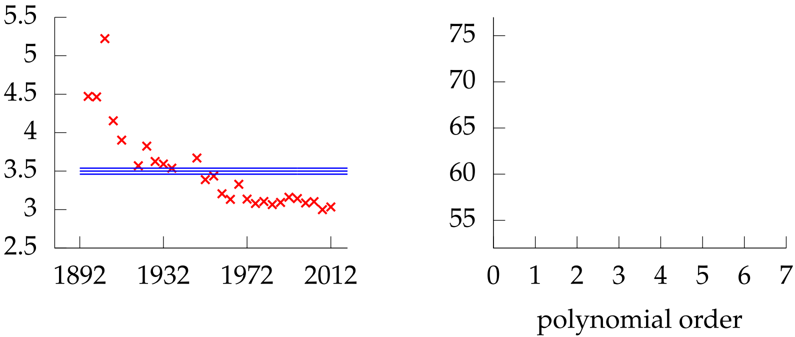
$$p(\mathbf{w}|\mathbf{y}, \mathbf{X})p(\mathbf{y}|\mathbf{X}) = p(\mathbf{y}|\mathbf{w}, \mathbf{X})p(\mathbf{w})$$

Marginal Likelihood

- Can compute the marginal likelihood as:

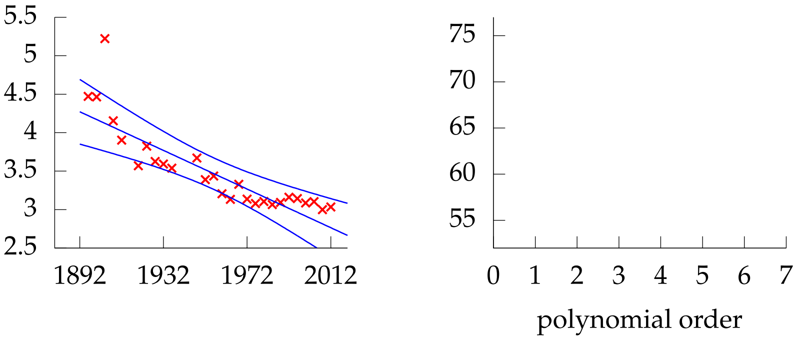
$$p(\mathbf{y}|\mathbf{X}, \alpha, \sigma) = \mathcal{N}\left(\mathbf{y}|\mathbf{0}, \alpha\mathbf{X}\mathbf{X}^\top + \sigma^2\mathbf{I}\right)$$

Polynomial Fits to Olympics Data



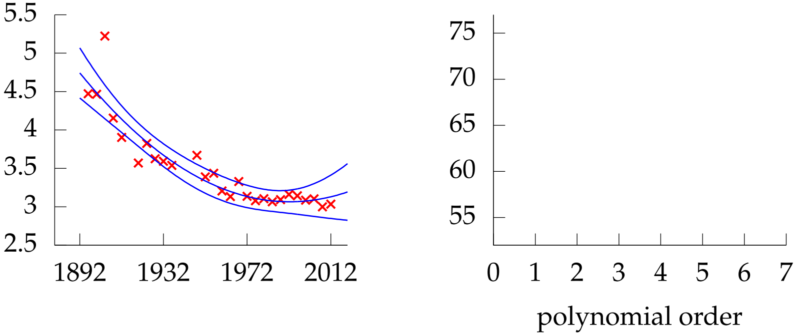
Left: fit to data, *Right*: marginal log likelihood. Polynomial order 0, model error 29.757, $\sigma^2 = 0.286$, $\sigma = 0.535$.

Polynomial Fits to Olympics Data



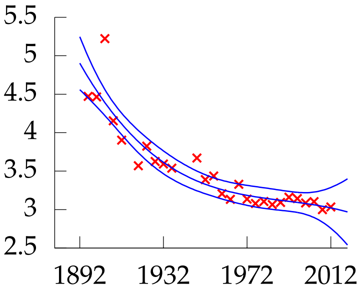
Left: fit to data, *Right*: marginal log likelihood. Polynomial order 1, model error 14.942, $\sigma^2 = 0.0749$, $\sigma = 0.274$.

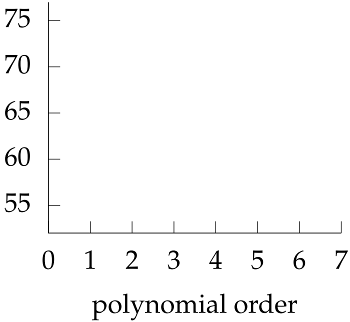
Polynomial Fits to Olympics Data



Left: fit to data, *Right*: marginal log likelihood. Polynomial order 2, model error 9.7206, $\sigma^2 = 0.0427$, $\sigma = 0.207$.

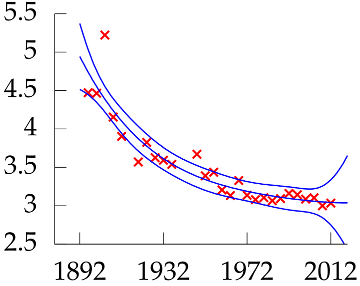
Polynomial Fits to Olympics Data

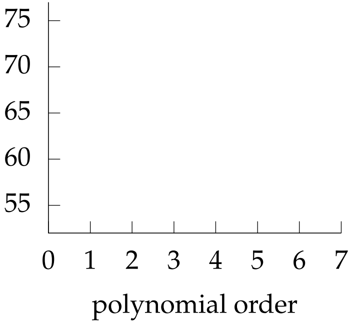




Left: fit to data, *Right*: marginal log likelihood. Polynomial order 3, model error 10.416, $\sigma^2 = 0.0402$, $\sigma = 0.200$.

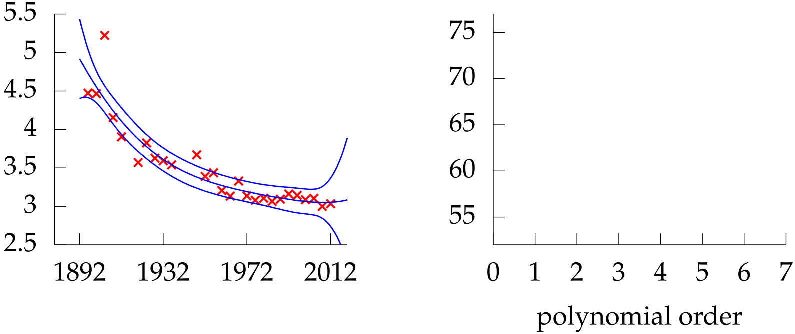
Polynomial Fits to Olympics Data





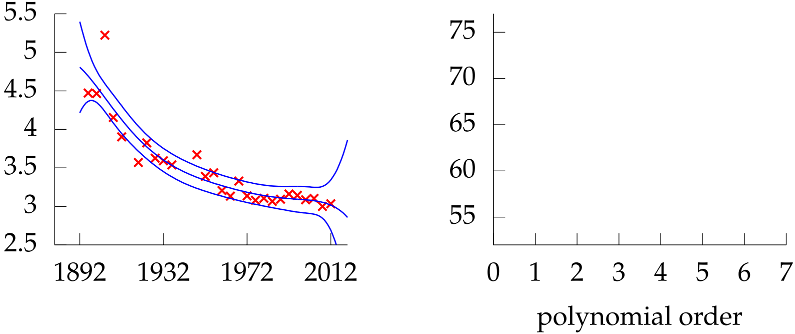
Left: fit to data, *Right*: marginal log likelihood. Polynomial order 4, model error 11.34, $\sigma^2 = 0.0401$, $\sigma = 0.200$.

Polynomial Fits to Olympics Data



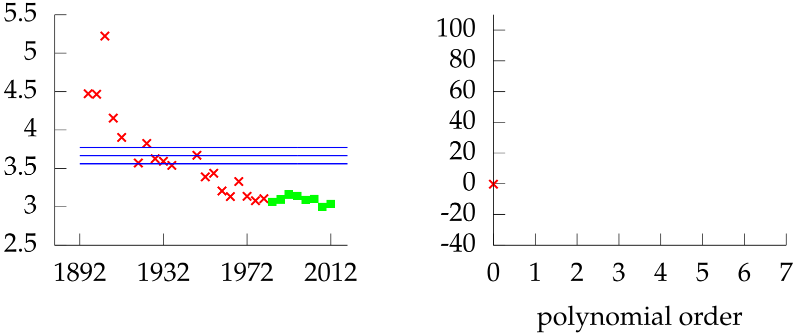
Left: fit to data, *Right*: marginal log likelihood. Polynomial order 5, model error 11.986, $\sigma^2 = 0.0399$, $\sigma = 0.200$.

Polynomial Fits to Olympics Data



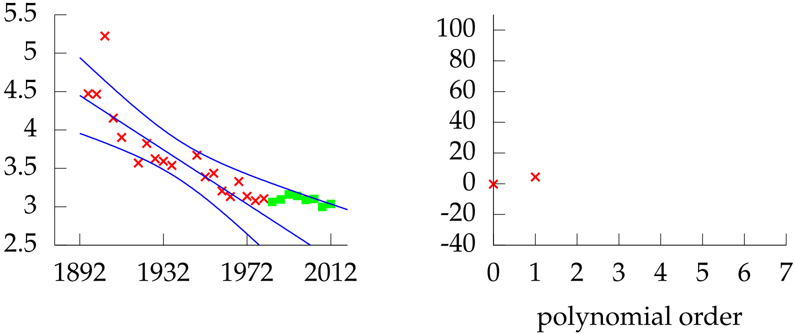
Left: fit to data, *Right*: marginal log likelihood. Polynomial order 6, model error 12.369, $\sigma^2 = 0.0384$, $\sigma = 0.196$.

Validation Set



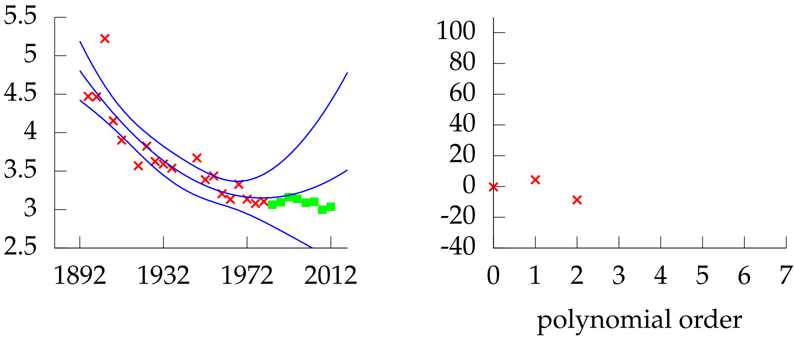
Left: fit to data, *Right*: model error. Polynomial order 0, training error 29.757, validation error -0.29243, $\sigma^2 = 0.302$, $\sigma = 0.550$.

Validation Set



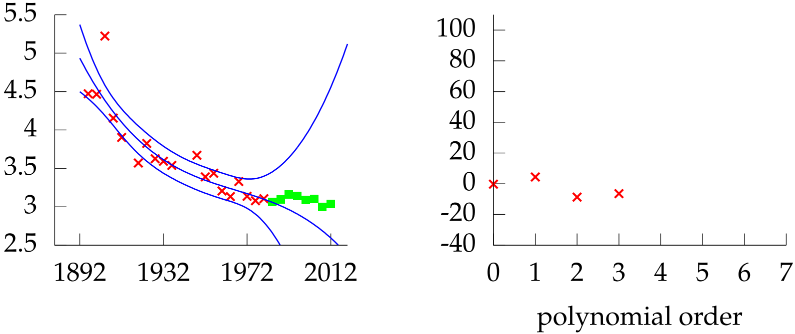
Left: fit to data, *Right*: model error. Polynomial order 1, training error 14.942, validation error 4.4027, $\sigma^2 = 0.0762$, $\sigma = 0.276$.

Validation Set



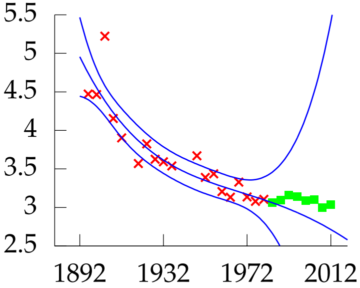
Left: fit to data, *Right*: model error. Polynomial order 2, training error 9.7206, validation error -8.6623, $\sigma^2 = 0.0580$, $\sigma = 0.241$.

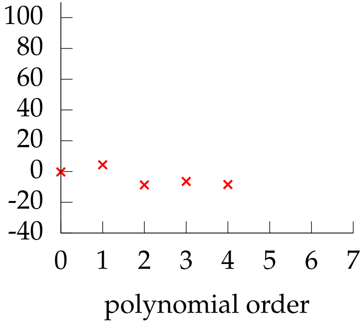
Validation Set



Left: fit to data, *Right*: model error. Polynomial order 3, training error 10.416, validation error -6.4726, $\sigma^2 = 0.0555$, $\sigma = 0.236$.

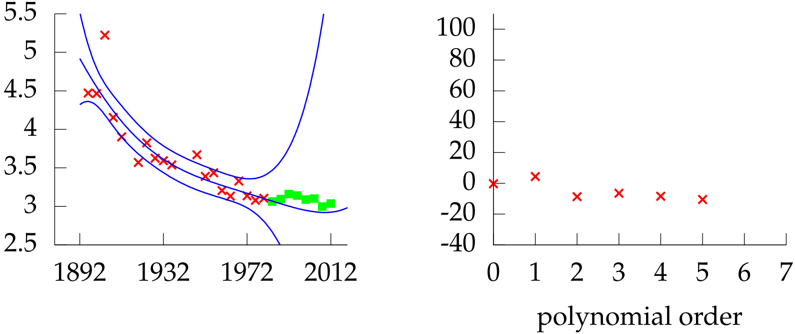
Validation Set





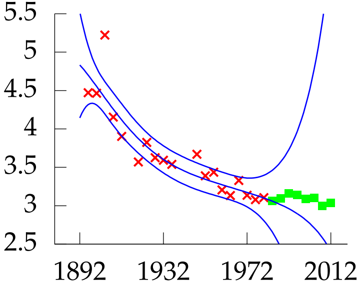
Left: fit to data, *Right*: model error. Polynomial order 4, training error 11.34, validation error -8.431, $\sigma^2 = 0.0555$, $\sigma = 0.236$.

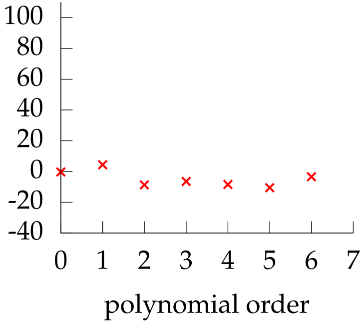
Validation Set



Left: fit to data, *Right*: model error. Polynomial order 5, training error 11.986, validation error -10.483, $\sigma^2 = 0.0551$, $\sigma = 0.235$.

Validation Set



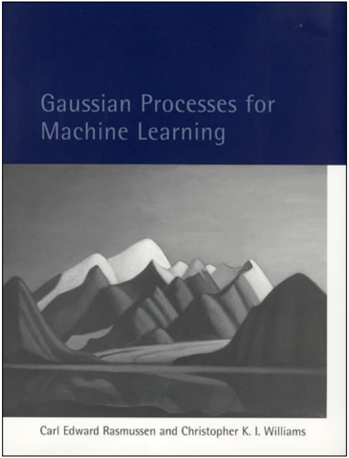


Left: fit to data, *Right*: model error. Polynomial order 6, training error 12.369, validation error -3.3823, $\sigma^2 = 0.0537$, $\sigma = 0.232$.

Reading

- Section 2.3 of Bishop up to top of pg 85 (multivariate Gaussians).
- Section 3.3 of Bishop up to 159 (pg 152–159).

Book



Rasmussen and Williams (2006)

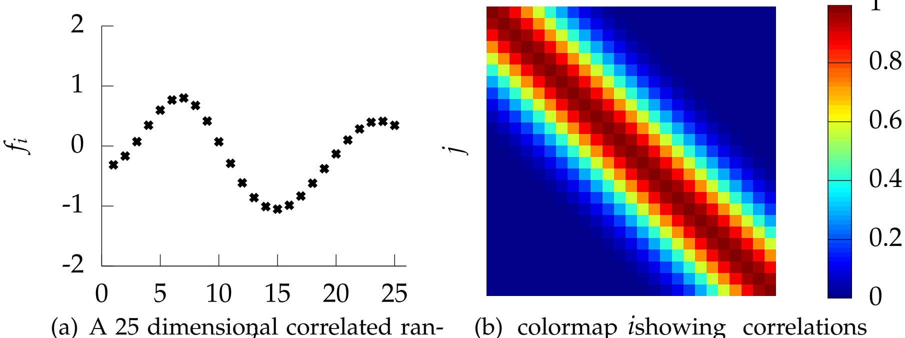
Outline

Sampling a Function

Multi-variate Gaussians

- We will consider a Gaussian with a particular structure of covariance matrix.
- Generate a single sample from this 25 dimensional Gaussian distribution, $\mathbf{f} = [f_1, f_2 \dots f_{25}]$.
- We will plot these points against their index.

Gaussian Distribution Sample

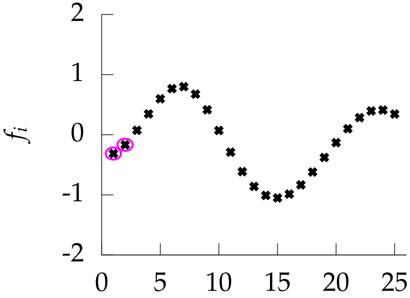


(a) A 25 dimensional correlated random variable (values ploted against index)

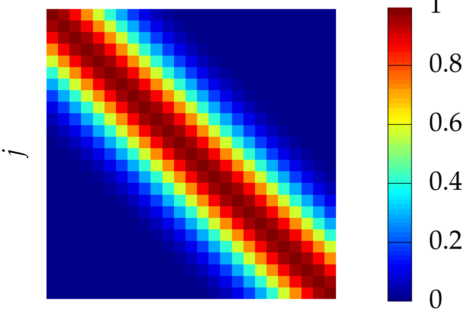
(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



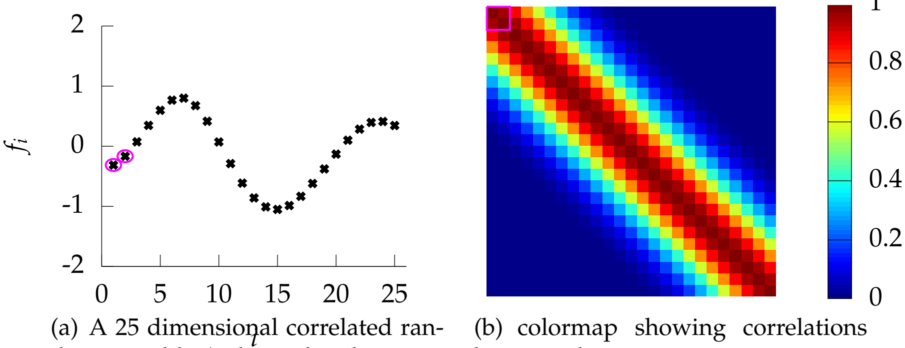
(a) A 25 dimensional correlated random variable (values ploted against index)



(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample

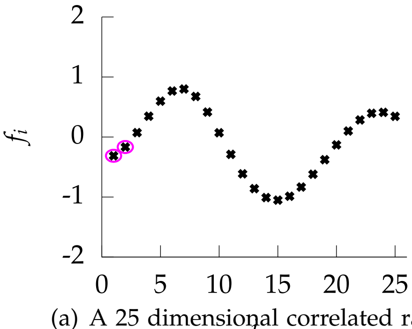


(a) A 25 dimensional correlated random variable (values ploted against index)

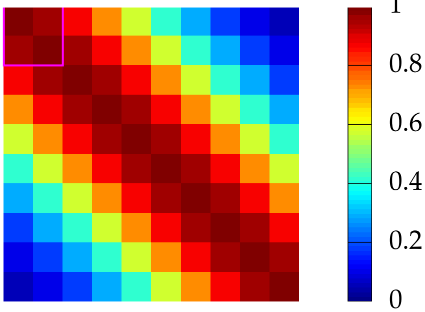
(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



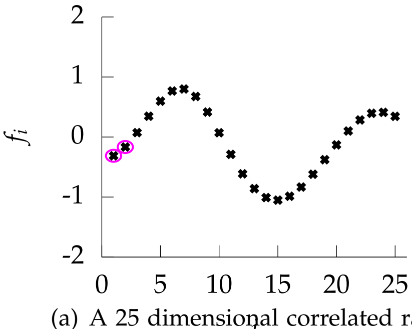
(a) A 25 dimensional correlated random variable (values ploted against index)



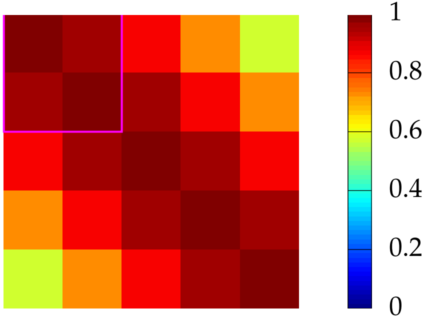
(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



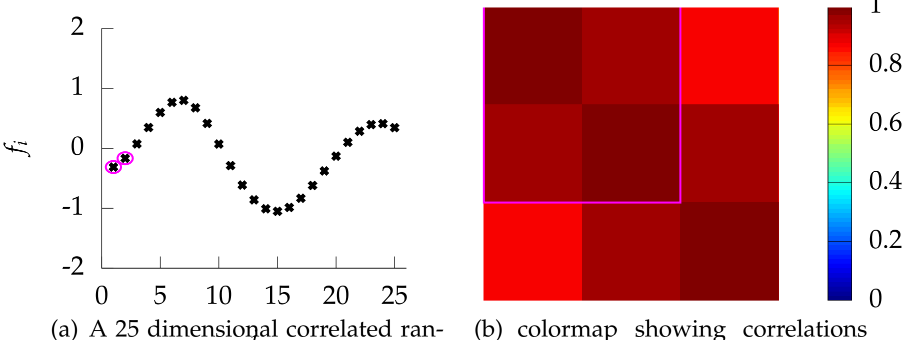
(a) A 25 dimensional correlated random variable (values ploted against index)



(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample

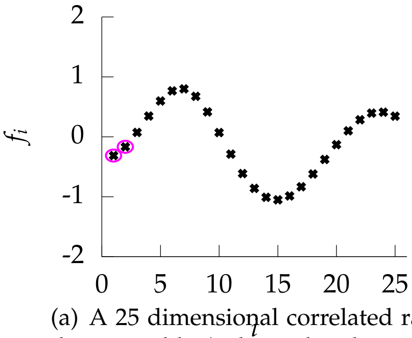


(a) A 25 dimensional correlated random variable (values ploted against index)

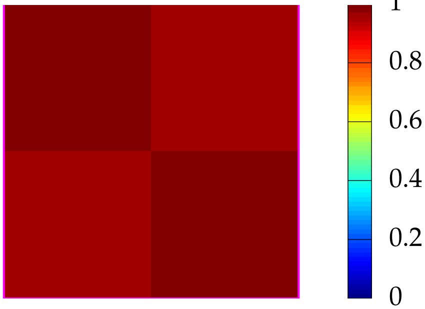
(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



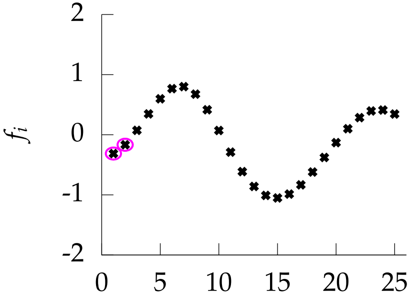
(a) A 25 dimensional correlated random variable (values ploted against index)



(b) colormap showing correlations between dimensions.

Figure: A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



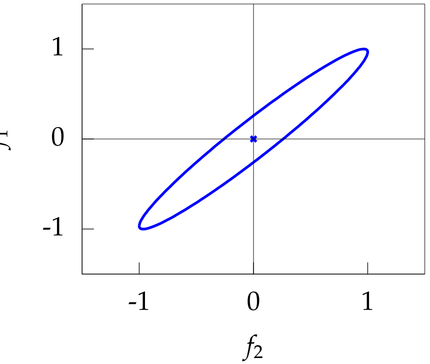
(a) A 25 dimensional correlated random variable (values ploted against index)

$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

(b) correlation between f_1 and f_2.

Figure: A sample from a 25 dimensional Gaussian distribution.

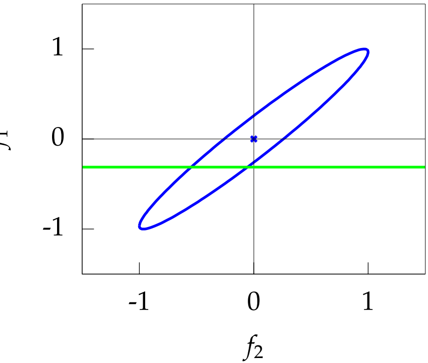
Prediction of f_2 from f_1



$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_2)$.

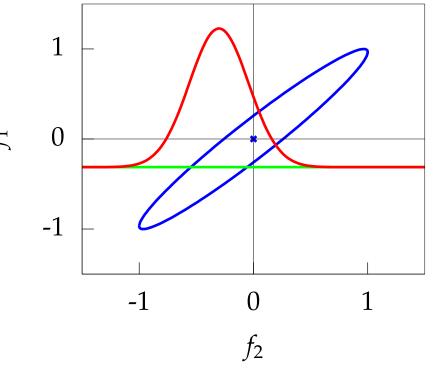
Prediction of f_2 from f_1



$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_2)$.
- We observe that $f_1 = -0.313$.

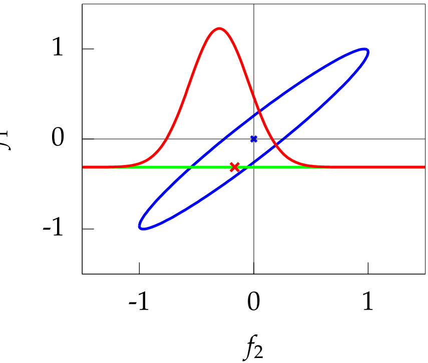
Prediction of f_2 from f_1



$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_2)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_2|f_1 = -0.313)$.

Prediction of f_2 from f_1



$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_2)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_2 | f_1 = -0.313)$.

Prediction with Correlated Gaussians

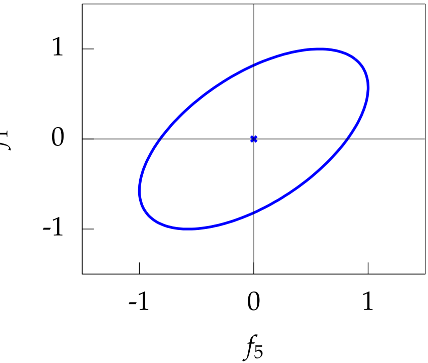
- Prediction of f_2 from f_1 requires *conditional density*.
- Conditional density is *also* Gaussian.

$$p(f_2|f_1) = \mathcal{N}\left(f_2|\frac{k_{1,2}}{k_{1,1}}f_1, k_{2,2} - \frac{k_{1,2}^2}{k_{1,1}}\right)$$

where covariance of joint density is given by

$$\mathbf{K} = \begin{bmatrix} k_{1,1} & k_{1,2} \\ k_{2,1} & k_{2,2} \end{bmatrix}$$

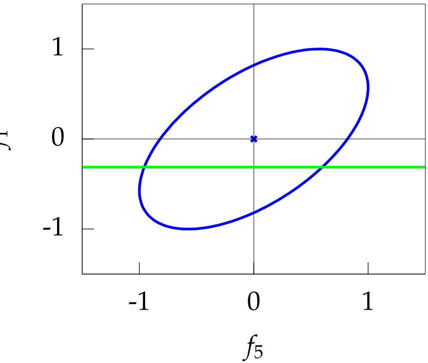
Prediction of f_5 from f_1



$$\begin{bmatrix} 1 & 0.57375 \\ 0.57375 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_5)$.

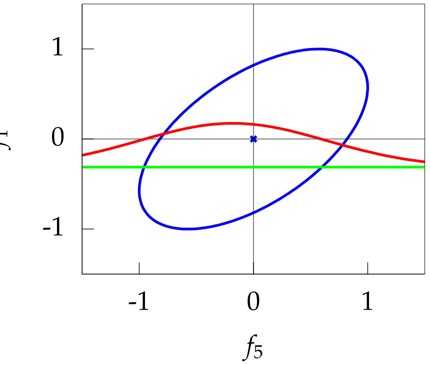
Prediction of f_5 from f_1



$$\begin{bmatrix} 1 & 0.57375 \\ 0.57375 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_5)$.
- We observe that $f_1 = -0.313$.

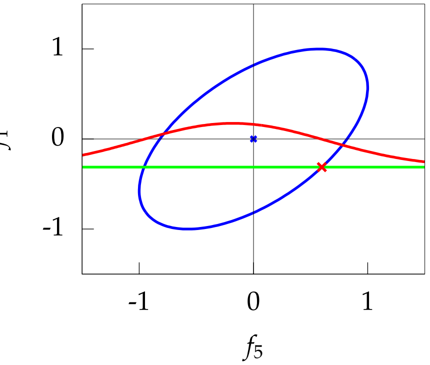
Prediction of f_5 from f_1



$$\begin{bmatrix} 1 & 0.57375 \\ 0.57375 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_5)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_5 | f_1 = -0.313)$.

Prediction of f_5 from f_1



$$\begin{bmatrix} 1 & 0.57375 \\ 0.57375 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_5)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_5 | f_1 = -0.313)$.

Prediction with Correlated Gaussians

- Prediction of $\mathbf{f}_*$ from $\mathbf{f}$ requires multivariate *conditional density*.
- Multivariate conditional density is *also* Gaussian.

$$p(\mathbf{f}_*|\mathbf{f}) = \mathcal{N}\left(\mathbf{f}_*|\mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{f}, \mathbf{K}_{*,*} - \mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{K}_{\mathbf{f},*}\right)$$

- Here covariance of joint density is given by

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{\mathbf{f},\mathbf{f}} & \mathbf{K}_{*,\mathbf{f}} \\ \mathbf{K}_{\mathbf{f},*} & \mathbf{K}_{*,*} \end{bmatrix}$$

Prediction with Correlated Gaussians

- Prediction of $\mathbf{f}_*$ from $\mathbf{f}$ requires multivariate *conditional density*.
- Multivariate conditional density is *also* Gaussian.

$$p(\mathbf{f}_*|\mathbf{f}) = \mathcal{N}\left(\mathbf{f}_*|\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

$$\boldsymbol{\mu} = \mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{f}$$

$$\boldsymbol{\Sigma} = \mathbf{K}_{*,*} - \mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{K}_{\mathbf{f},*}$$

- Here covariance of joint density is given by

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{\mathbf{f},\mathbf{f}} & \mathbf{K}_{*,\mathbf{f}} \\ \mathbf{K}_{\mathbf{f},*} & \mathbf{K}_{*,*} \end{bmatrix}$$

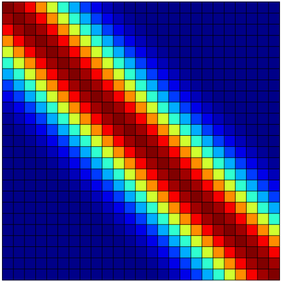
Covariance Functions

Where did this covariance matrix come from?

Exponentiated Quadratic Kernel Function (RBF, Squared Exponential, Gaussian)

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function $\mathbf{x}$.
- For the example above it was based on Euclidean distance.
- The covariance function is also know as a kernel.



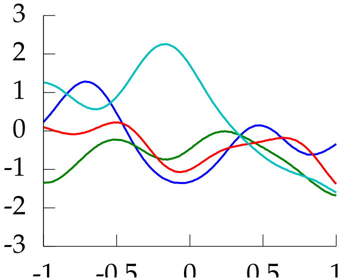
Covariance Functions

Where did this covariance matrix come from?

Exponentiated Quadratic Kernel Function (RBF, Squared Exponential, Gaussian)

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function $\mathbf{x}$.
- For the example above it was based on Euclidean distance.
- The covariance function is also know as a kernel.



Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_1 = -3.0, x_1 = -3.0$$

$$k_{1,1} = 1.00 \times \exp\left(-\frac{(-3.0 - -3.0)^2}{2 \times 2.00^2}\right)$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_1 = -3.0, x_1 = -3.0$$

$$k_{1,1} = 1.00 \times \exp\left(-\frac{(-3.0 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & & \\ & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20, x_1 = -3.0$$

$$k_{2,1} = 1.00 \times \exp\left(-\frac{(1.20 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & & \\ & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20,\ x_1 = -3.0$$

$$k_{2,1} = 1.00 \times \exp\left(-\frac{(1.20 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & & \\ 0.110 & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20, x_1 = -3.0$$

$$k_{2,1} = 1.00 \times \exp\left(-\frac{(1.20 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20, x_2 = 1.20$$

$$k_{2,2} = 1.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_2 = 1.20, x_2 = 1.20$

$$k_{2,2} = 1.00 \times \exp\left(-\frac{(1.20-1.20)^2}{2\times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & 1.00 & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_1 = -3.0$$

$$k_{3,1} = 1.00 \times \exp\left(-\frac{(1.40 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & 1.00 & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\, x_1 = -3.0$$

$$k_{3,1} = 1.00 \times \exp\left(-\frac{(1.40 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & 1.00 & \\ 0.0889 & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_1 = -3.0$$

$$k_{3,1} = 1.00 \times \exp\left(-\frac{(1.40 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & \\ 0.0889 & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\, x_2 = 1.20$$

$$k_{3,2} = 1.00 \times \exp\left(-\frac{(1.40 - 1.20)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & \\ 0.0889 & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\ x_2 = 1.20$$

$$k_{3,2} = 1.00 \times \exp\left(-\frac{(1.40-1.20)^2}{2\times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & \\ 0.0889 & \boxed{0.995} & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_2 = 1.20$$

$$k_{3,2} = 1.00 \times \exp\left(-\frac{(1.40-1.20)^2}{2\times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & 0.995 \\ 0.0889 & 0.995 & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\, x_3 = 1.40$$

$$k_{3,3} = 1.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & 0.995 \\ 0.0889 & 0.995 & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_3 = 1.40$$

$$k_{3,3} = 1.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & 0.995 \\ 0.0889 & 0.995 & \boxed{1.00} \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

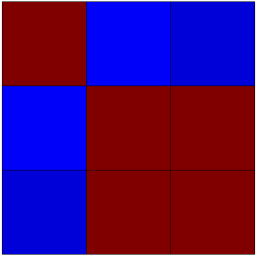
Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_3 = 1.40$, $x_3 = 1.40$

$$k_{3,3} = 1.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 2.00^2}\right)$$



$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_1 = -3, x_1 = -3$$

$$k_{1,1} = 1.0 \times \exp\left(-\frac{(-3 - -3)^2}{2 \times 2.0^2}\right)$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_1 = -3, x_1 = -3$

$$k_{1,1} = 1.0 \times \exp\left(-\frac{(-3 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.2,\ x_1 = -3$$

$$k_{2,1} = 1.0 \times \exp\left(-\frac{(1.2 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.2, x_1 = -3$$

$$k_{2,1} = 1.0 \times \exp\left(-\frac{(1.2 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & & & \\ \boxed{0.11} & & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.2, x_1 = -3$$

$$k_{2,1} = 1.0 \times \exp\left(-\frac{(1.2 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.2, x_2 = 1.2$$

$$k_{2,2} = 1.0 \times \exp\left(-\frac{(1.2-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.2,\ x_2 = 1.2$$

$$k_{2,2} = 1.0 \times \exp\left(-\frac{(1.2-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & \boxed{1.0} & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4, x_1 = -3$$

$$k_{3,1} = 1.0 \times \exp\left(-\frac{(1.4 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & 1.0 & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_3 = 1.4$, $x_1 = -3$

$$k_{3,1} = 1.0 \times \exp\left(-\frac{(1.4 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & 1.0 & & \\ 0.089 & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4,\, x_1 = -3$$

$$k_{3,1} = 1.0 \times \exp\left(-\frac{(1.4 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & & \\ 0.089 & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4,\ x_2 = 1.2$$

$$k_{3,2} = 1.0 \times \exp\left(-\frac{(1.4-1.2)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & & \\ 0.089 & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_3 = 1.4$, $x_2 = 1.2$

$$k_{3,2} = 1.0 \times \exp\left(-\frac{(1.4-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & & \\ 0.089 & \boxed{1.0} & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4,\ x_2 = 1.2$$

$$k_{3,2} = 1.0 \times \exp\left(-\frac{(1.4-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4,\ x_3 = 1.4$$

$$k_{3,3} = 1.0 \times \exp\left(-\frac{(1.4-1.4)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4, x_3 = 1.4$$

$$k_{3,3} = 1.0 \times \exp\left(-\frac{(1.4-1.4)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 \\ 0.11 & 1.0 & 1.0 \\ 0.089 & 1.0 & \boxed{1.0} \\ & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_1 = -3$$

$$k_{4,1} = 1.0 \times \exp\left(-\frac{(2.0 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_1 = -3$$

$$k_{4,1} = 1.0 \times \exp\left(-\frac{(2.0 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_1 = -3$$

$$k_{4,1} = 1.0 \times \exp\left(-\frac{(2.0 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0,\ x_2 = 1.2$$

$$k_{4,2} = 1.0 \times \exp\left(-\frac{(2.0-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_4 = 2.0, x_2 = 1.2$

$$k_{4,2} = 1.0 \times \exp\left(-\frac{(2.0-1.2)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & \boxed{0.92} & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_2 = 1.2$$

$$k_{4,2} = 1.0 \times \exp\left(-\frac{(2.0-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & 0.92 & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_3 = 1.4$$

$$k_{4,3} = 1.0 \times \exp\left(-\frac{(2.0-1.4)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & 0.92 & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_4 = 2.0, x_3 = 1.4$

$$k_{4,3} = 1.0 \times \exp\left(-\frac{(2.0-1.4)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & 0.92 & \boxed{0.96} & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_3 = 1.4$$

$$k_{4,3} = 1.0 \times \exp\left(-\frac{(2.0-1.4)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & 0.96 \\ 0.044 & 0.92 & 0.96 & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_4 = 2.0$$

$$k_{4,4} = 1.0 \times \exp\left(-\frac{(2.0 - 2.0)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & 0.96 \\ 0.044 & 0.92 & 0.96 & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_4 = 2.0$$

$$k_{4,4} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & 0.96 \\ 0.044 & 0.92 & 0.96 & \boxed{1.0} \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

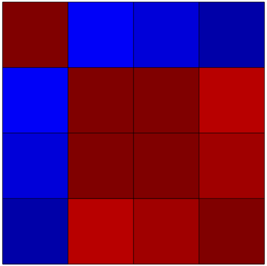
Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_4 = 2.0, x_4 = 2.0$

$$k_{4,4} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2\times 2.0^2}\right)$$



$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_1 = -3.0, x_1 = -3.0$$

$$k_{1,1} = 4.00 \times \exp\left(-\frac{(-3.0 - -3.0)^2}{2 \times 5.00^2}\right)$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_1 = -3.0, x_1 = -3.0$$

$$k_{1,1} = 4.00 \times \exp\left(-\frac{(-3.0 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & & \\ & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k(x_i, x_j) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_2 = 1.20$, $x_1 = -3.0$

$$k_{2,1} = 4.00 \times \exp\left(-\frac{(1.20 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & & \\ & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20, x_1 = -3.0$$

$$k_{2,1} = 4.00 \times \exp\left(-\frac{(1.20 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & & \\ \boxed{2.81} & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20, x_1 = -3.0$$

$$k_{2,1} = 4.00 \times \exp\left(-\frac{(1.20 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & \\ 2.81 & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20, x_2 = 1.20$$

$$k_{2,2} = 4.00 \times \exp\left(-\frac{(1.20-1.20)^2}{2\times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & \\ 2.81 & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20, x_2 = 1.20$$

$$k_{2,2} = 4.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 \\ 2.81 & \boxed{4.00} \\ & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_1 = -3.0$$

$$k_{3,1} = 4.00 \times \exp\left(-\frac{(1.40 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 \\ 2.81 & 4.00 \\ & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_1 = -3.0$$

$$k_{3,1} = 4.00 \times \exp\left(-\frac{(1.40 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & \\ 2.81 & 4.00 & \\ \boxed{2.72} & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_1 = -3.0$$

$$k_{3,1} = 4.00 \times \exp\left(-\frac{(1.40 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & \\ 2.72 & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\ x_2 = 1.20$$

$$k_{3,2} = 4.00 \times \exp\left(-\frac{(1.40 - 1.20)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & \\ 2.72 & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\ x_2 = 1.20$$

$$k_{3,2} = 4.00 \times \exp\left(-\frac{(1.40-1.20)^2}{2\times 5.00^2}\right)$$

4.00	2.81	2.72
2.81	4.00	
2.72	4.00	

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_2 = 1.20$$

$$k_{3,2} = 4.00 \times \exp\left(-\frac{(1.40-1.20)^2}{2\times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & 4.00 \\ 2.72 & 4.00 & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_3 = 1.40$$

$$k_{3,3} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & 4.00 \\ 2.72 & 4.00 & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_3 = 1.40$$

$$k_{3,3} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & 4.00 \\ 2.72 & 4.00 & \boxed{4.00} \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

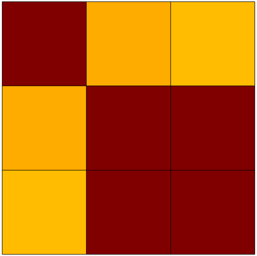
Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_3 = 1.40$, $x_3 = 1.40$

$$k_{3,3} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$



$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Outline

Basis Function Form

Radial basis functions commonly have the form

$$\phi_k(\mathbf{x}_i) = \exp\left(-\frac{\left|\mathbf{x}_i - \boldsymbol{\mu}_k\right|^2}{2\ell^2}\right).$$

- Basis function maps data into a "feature space" in which a linear sum is a non linear function.

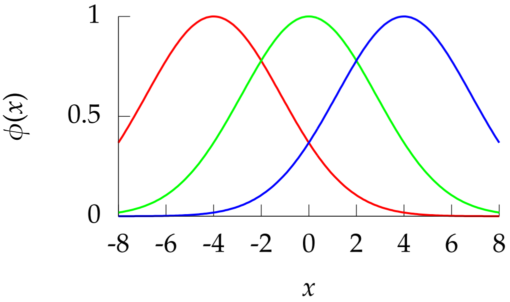


Figure: A set of radial basis functions with width $\ell = 2$ and location parameters $\boldsymbol{\mu} = [-4 \;\; 0 \;\; 4]^\top$.

Basis Function Representations

- Represent a function by a linear sum over a basis,

$$f(\mathbf{x}_{i,:};\mathbf{w}) = \sum_{k=1}^{m} w_k \phi_k(\mathbf{x}_{i,:}), \tag{1}$$

- Here: m basis functions and $\phi_k(\cdot)$ is kth basis function and

$$\mathbf{w} = [w_1, \ldots, w_m]^\top .$$

- For standard linear model: $\phi_k(\mathbf{x}_{i,:}) = x_{i,k}$.

Random Functions

Functions derived using:

$$f(x) = \sum_{k=1}^{m} w_k \phi_k(x),$$

where $\mathbf{W}$ is sampled from a Gaussian density,

$$w_k \sim \mathcal{N}(0, \alpha).$$

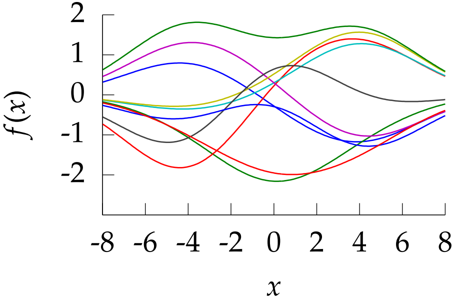


Figure: Functions sampled using the basis set from figure 2. Each line is a separate sample, generated by a weighted sum of the basis set. The weights, $\mathbf{w}$ are sampled from a Gaussian density with variance $\alpha = 1$.

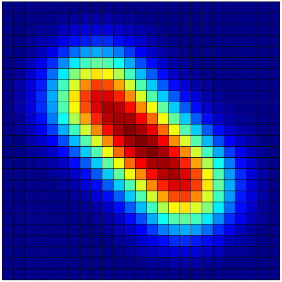
Covariance Functions

RBF Basis Functions

$$k(\mathbf{x}, \mathbf{x}') = \alpha \boldsymbol{\phi}(\mathbf{x})^\top \boldsymbol{\phi}(\mathbf{x}')$$

$$\phi_i(x) = \exp\left(-\frac{\left\|x - \mu_i\right\|_2^2}{\ell^2}\right)$$

$$\boldsymbol{\mu} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$



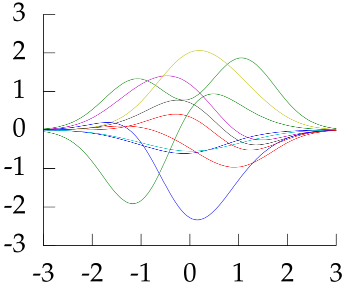
Covariance Functions

RBF Basis Functions

$$k(\mathbf{x}, \mathbf{x}') = \alpha \boldsymbol{\phi}(\mathbf{x})^\top \boldsymbol{\phi}(\mathbf{x}')$$

$$\phi_i(x) = \exp\left(-\frac{\left\|x - \mu_i\right\|_2^2}{\ell^2}\right)$$

$$\boldsymbol{\mu} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$



Direct Construction of Covariance Matrix

- Use matrix notation to write function,

$$f(\mathbf{x}_i; \mathbf{w}) = \sum_{k=1}^{m} w_k \phi_k(\mathbf{x}_i)$$

Direct Construction of Covariance Matrix

- Use matrix notation to write function,

$$f(\mathbf{x}_i; \mathbf{w}) = \sum_{k=1}^{m} w_k \phi_k(\mathbf{x}_i)$$

computed at training data gives a vector

$$\mathbf{f} = \mathbf{\Phi} \mathbf{w}.$$

Direct Construction of Covariance Matrix

- Use matrix notation to write function,

$$f(\mathbf{x}_i; \mathbf{w}) = \sum_{k=1}^{m} w_k \phi_k(\mathbf{x}_i)$$

computed at training data gives a vector

$$\mathbf{f} = \mathbf{\Phi}\mathbf{w}.$$

w and **f** are only related by an *inner product*.

Direct Construction of Covariance Matrix

- Use matrix notation to write function,

$$f(\mathbf{x}_i; \mathbf{w}) = \sum_{k=1}^{m} w_k \phi_k(\mathbf{x}_i)$$

computed at training data gives a vector

$$\mathbf{f} = \boldsymbol{\Phi}\mathbf{w}.$$

$\mathbf{w}$ and $\mathbf{f}$ are only related by an *inner product*.

$\boldsymbol{\Phi} \in \Re^{n \times p}$ is a *design matrix*

Direct Construction of Covariance Matrix

- Use matrix notation to write function,

$$f(\mathbf{x}_i; \mathbf{w}) = \sum_{k=1}^{m} w_k \phi_k(\mathbf{x}_i)$$

computed at training data gives a vector

$$\mathbf{f} = \boldsymbol{\Phi}\mathbf{w}.$$

w and **f** are only related by an *inner product*.

$\boldsymbol{\Phi} \in \Re^{n \times p}$ is a *design matrix*

$\boldsymbol{\Phi}$ is fixed and non-stochastic for a given training set.

Direct Construction of Covariance Matrix

- Use matrix notation to write function,

$$f(\mathbf{x}_i; \mathbf{w}) = \sum_{k=1}^{m} w_k \phi_k(\mathbf{x}_i)$$

computed at training data gives a vector

$$\mathbf{f} = \boldsymbol{\Phi}\mathbf{w}.$$

$\mathbf{w}$ and $\mathbf{f}$ are only related by an *inner product*.

$\boldsymbol{\Phi} \in \Re^{n\times p}$ is a *design matrix*

$\boldsymbol{\Phi}$ is fixed and non-stochastic for a given training set.

$\mathbf{f}$ is Gaussian distributed.

Expectations

- We have

$$\langle \mathbf{f} \rangle = \mathbf{\Phi} \langle \mathbf{w} \rangle .$$

We use $\langle \cdot \rangle$ to denote expectations under prior distributions.

Expectations

- We have

$$\langle \mathbf{f} \rangle = \mathbf{\Phi} \langle \mathbf{w} \rangle .$$

- Prior mean of $\mathbf{w}$ was zero giving

$$\langle \mathbf{f} \rangle = \mathbf{0}.$$

We use $\langle \cdot \rangle$ to denote expectations under prior distributions.

Expectations

- We have
$$\langle \mathbf{f} \rangle = \mathbf{\Phi} \langle \mathbf{w} \rangle .$$
- Prior mean of $\mathbf{w}$ was zero giving
$$\langle \mathbf{f} \rangle = \mathbf{0}.$$
- Prior covariance of $\mathbf{f}$ is
$$\mathbf{K} = \left\langle \mathbf{f}\mathbf{f}^\top \right\rangle - \langle \mathbf{f} \rangle \langle \mathbf{f} \rangle^\top$$

We use $\langle \cdot \rangle$ to denote expectations under prior distributions.

Expectations

- We have
$$\langle \mathbf{f} \rangle = \mathbf{\Phi} \langle \mathbf{w} \rangle .$$
- Prior mean of $\mathbf{w}$ was zero giving
$$\langle \mathbf{f} \rangle = \mathbf{0}.$$
- Prior covariance of $\mathbf{f}$ is
$$\mathbf{K} = \left\langle \mathbf{f}\mathbf{f}^\top \right\rangle - \langle \mathbf{f} \rangle \langle \mathbf{f} \rangle^\top$$
$$\left\langle \mathbf{f}\mathbf{f}^\top \right\rangle = \mathbf{\Phi} \left\langle \mathbf{w}\mathbf{w}^\top \right\rangle \mathbf{\Phi}^\top,$$
giving
$$\mathbf{K} = \gamma' \mathbf{\Phi}\mathbf{\Phi}^\top.$$

We use $\langle \cdot \rangle$ to denote expectations under prior distributions.

Covariance between Two Points

- The prior covariance between two points $\mathbf{x}_i$ and $\mathbf{x}_j$ is

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \phi_{:}\left(\mathbf{x}_i\right)^\top \phi_{:}\left(\mathbf{x}_j\right),$$

Covariance between Two Points

- The prior covariance between two points $\mathbf{x}_i$ and $\mathbf{x}_j$ is

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \phi_:\left(\mathbf{x}_i\right)^\top \phi_:\left(\mathbf{x}_j\right),$$

or in sum notation

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \gamma' \sum_{\ell}^{m} \phi_\ell\left(\mathbf{x}_i\right) \phi_\ell\left(\mathbf{x}_j\right)$$

Covariance between Two Points

- The prior covariance between two points $\mathbf{x}_i$ and $\mathbf{x}_j$ is

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \phi_:\left(\mathbf{x}_i\right)^\top \phi_:\left(\mathbf{x}_j\right),$$

  or in sum notation

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \gamma' \sum_{\ell}^{m} \phi_\ell\left(\mathbf{x}_i\right) \phi_\ell\left(\mathbf{x}_j\right)$$

- For the radial basis used this gives

Covariance between Two Points

- The prior covariance between two points $\mathbf{x}_i$ and $\mathbf{x}_j$ is

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \phi_{:}\left(\mathbf{x}_i\right)^{\top} \phi_{:}\left(\mathbf{x}_j\right),$$

  or in sum notation

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \gamma' \sum_{\ell}^{m} \phi_{\ell}\left(\mathbf{x}_i\right) \phi_{\ell}\left(\mathbf{x}_j\right)$$

- For the radial basis used this gives

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \gamma' \sum_{k=1}^{m} \exp\left(-\frac{\left|\mathbf{x}_i - \boldsymbol{\mu}_k\right|^2 + \left|\mathbf{x}_j - \boldsymbol{\mu}_k\right|^2}{2\ell^2}\right).$$

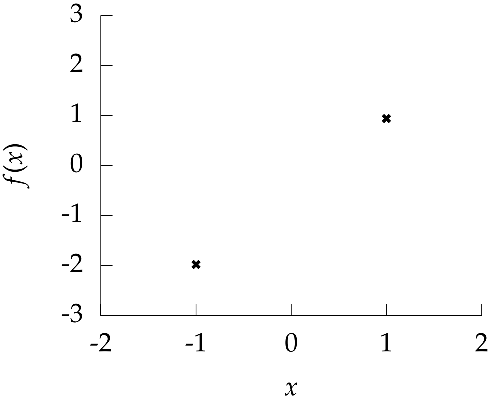


Figure: Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

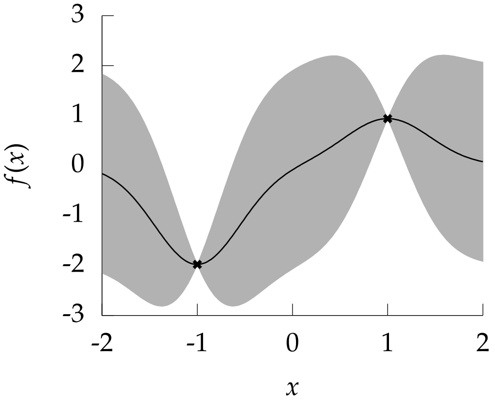


Figure: Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

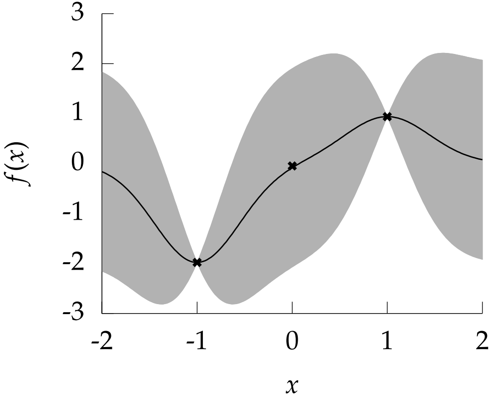


Figure: Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

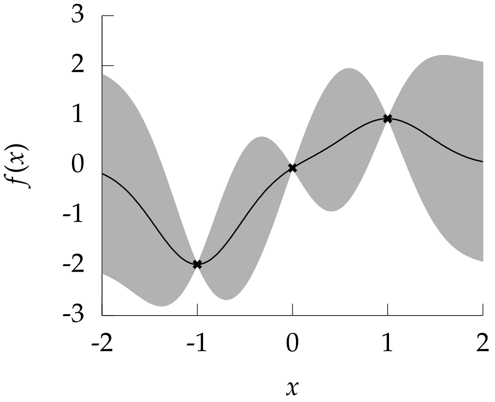


Figure: Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

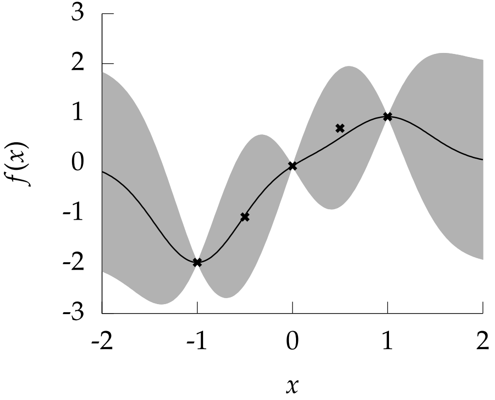


Figure: Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

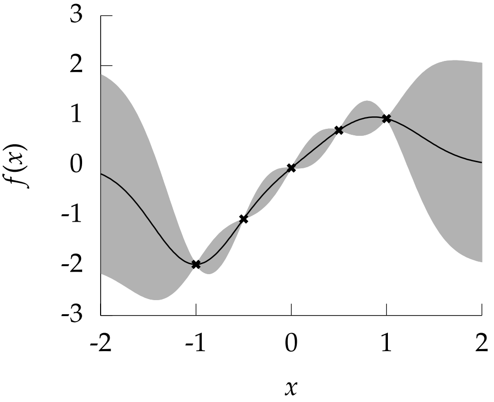


Figure: Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

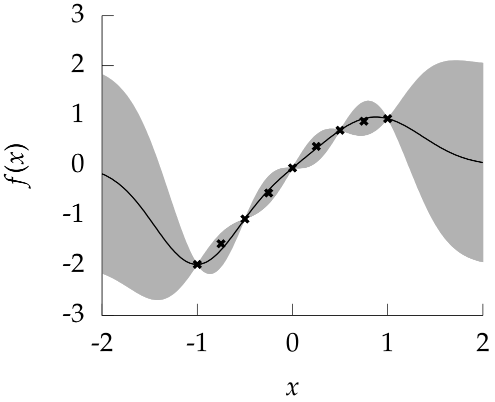


Figure: Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

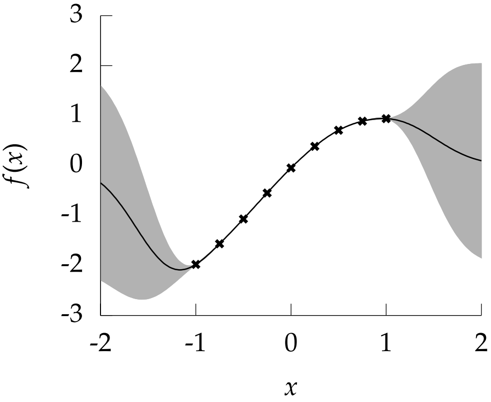


Figure: Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Regression

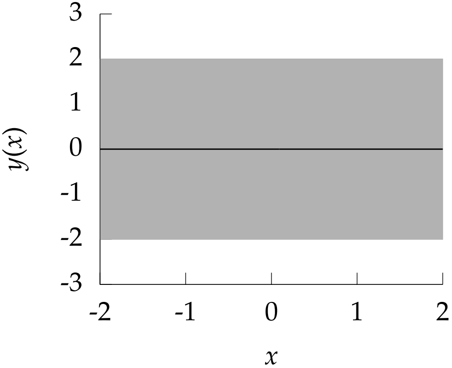


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

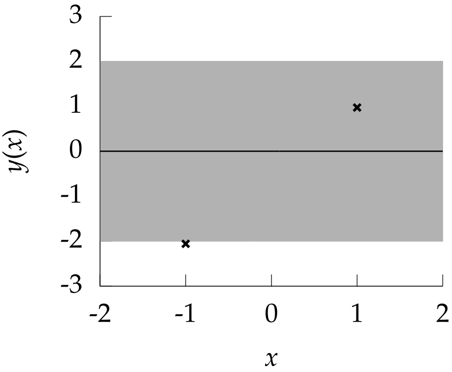


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

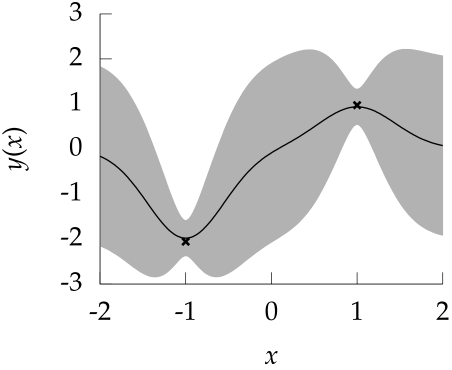


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

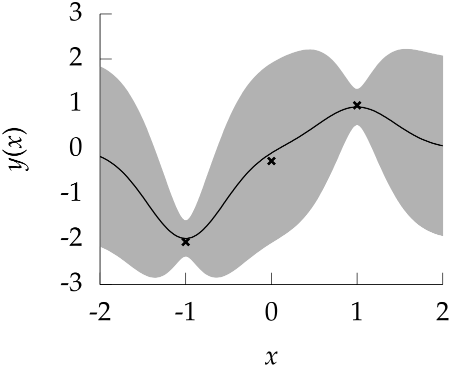


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

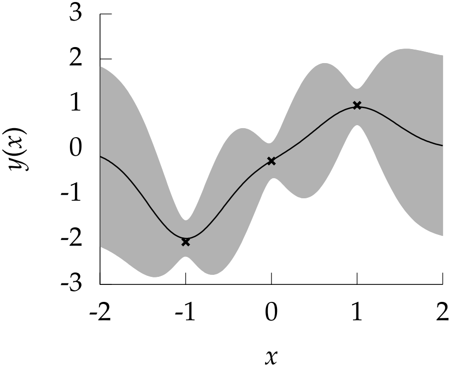


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

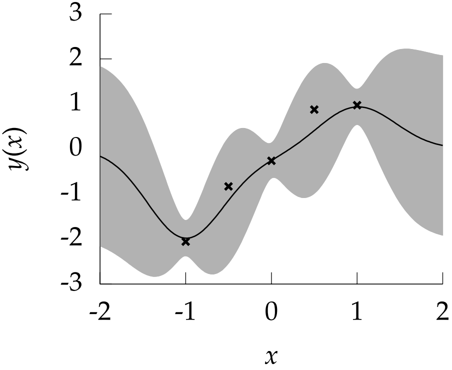


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

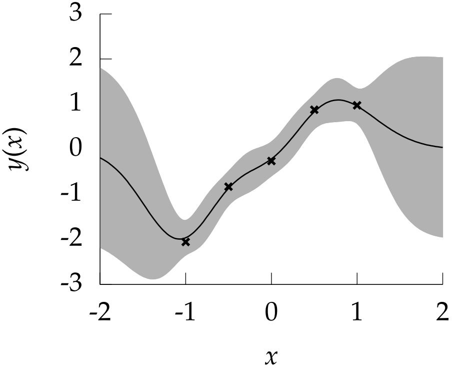


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

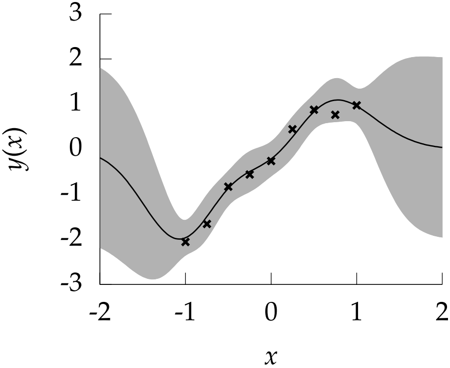


Figure: Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

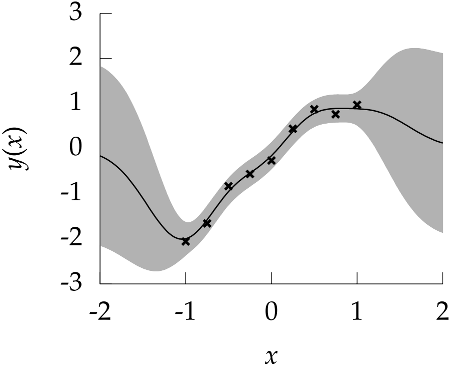


Figure: Examples include WiFi localization, C14 callibration curve.

Learning Covariance Parameters

Can we determine covariance parameters from the data?

$$\mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{K}|}\exp\left(-\frac{\mathbf{y}^\top\mathbf{K}^{-1}\mathbf{y}}{2}\right)$$

The parameters are *inside* the covariance function (matrix).

$$k_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta})$$

Learning Covariance Parameters

Can we determine covariance parameters from the data?

$$\mathcal{N}\left(\mathbf{y}|\mathbf{0},\mathbf{K}\right) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{K}|}\exp\left(-\frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}\right)$$

The parameters are *inside* the covariance function (matrix).

$$k_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta})$$

Learning Covariance Parameters

Can we determine covariance parameters from the data?

$$\log \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}) = -\frac{1}{2}\log|\mathbf{K}| - \frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2} - \frac{n}{2}\log 2\pi$$

The parameters are *inside* the covariance function (matrix).

$$k_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta})$$

Learning Covariance Parameters

Can we determine covariance parameters from the data?

$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

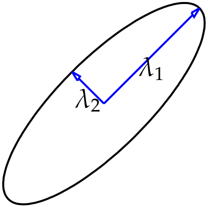
The parameters are *inside* the covariance function (matrix).

$$k_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta})$$

Eigendecomposition of Covariance

A useful decomposition for understanding the objective function.

$$\mathbf{K} = \mathbf{R}\boldsymbol{\Lambda}^2\mathbf{R}^\top$$

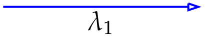


Diagonal of $\boldsymbol{\Lambda}$ represents distance along axes.
$\mathbf{R}$ gives a rotation of these axes.

where $\boldsymbol{\Lambda}$ is a *diagonal* matrix and $\mathbf{R}^\top\mathbf{R} = \mathbf{I}$.

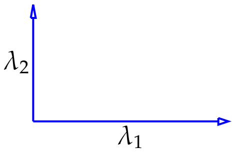
Capacity control: $\log |\mathbf{K}|$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



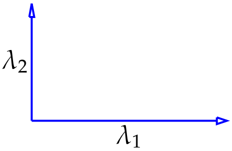
Capacity control: $\log |\mathbf{K}|$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



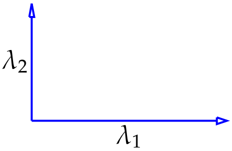
Capacity control: $\log |\mathbf{K}|$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



Capacity control: $\log |\mathbf{K}|$

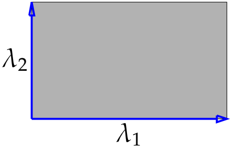
$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$|\mathbf{\Lambda}| = \lambda_1 \lambda_2$$

Capacity control: $\log |\mathbf{K}|$

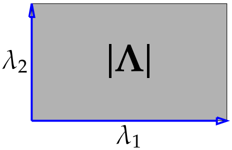
$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$|\boldsymbol{\Lambda}| = \lambda_1 \lambda_2$$

Capacity control: $\log |\mathbf{K}|$

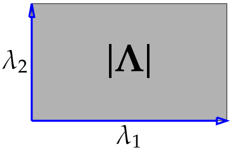
$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$|\mathbf{\Lambda}| = \lambda_1 \lambda_2$$

Capacity control: $\log |\mathbf{K}|$

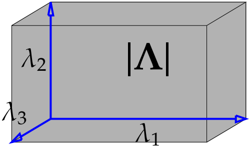
$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$



$$|\mathbf{\Lambda}| = \lambda_1 \lambda_2$$

Capacity control: $\log |\mathbf{K}|$

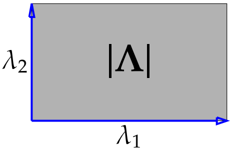
$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$



$$|\boldsymbol{\Lambda}| = \lambda_1 \lambda_2 \lambda_3$$

Capacity control: $\log |\mathbf{K}|$

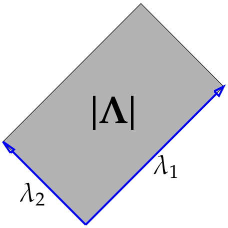
$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$|\mathbf{\Lambda}| = \lambda_1 \lambda_2$$

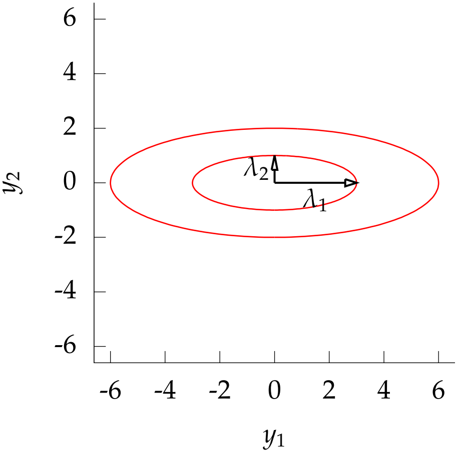
Capacity control: $\log |\mathbf{K}|$

$$\mathbf{R}\boldsymbol{\Lambda} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix}$$

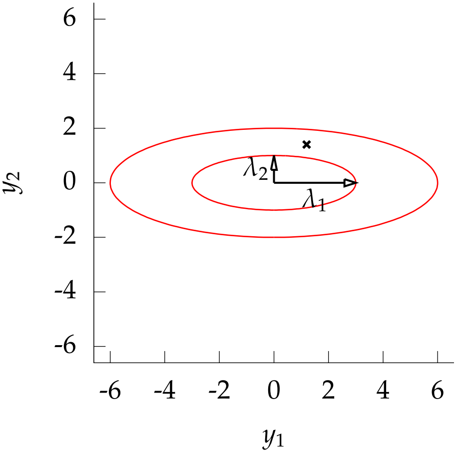


$$|\mathbf{R}\boldsymbol{\Lambda}| = \lambda_1 \lambda_2$$

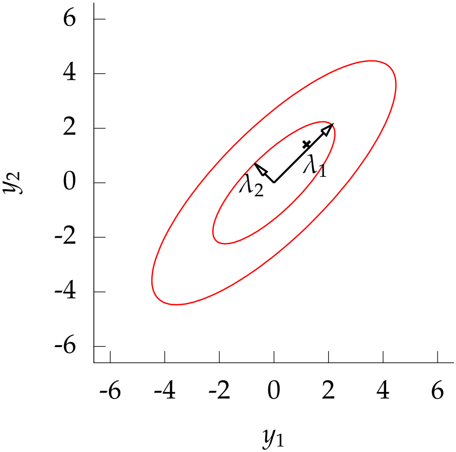
Data Fit: $\frac{\mathbf{y}^{-1}\mathbf{K}^{-1}\mathbf{y}}{2}$



Data Fit: $\frac{\mathbf{y}^{-1}\mathbf{K}^{-1}\mathbf{y}}{2}$

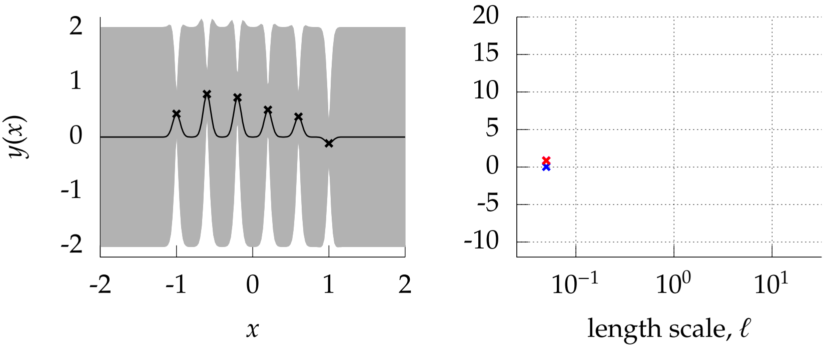


Data Fit: $\frac{\mathbf{y}^{-1}\mathbf{K}^{-1}\mathbf{y}}{2}$



Learning Covariance Parameters

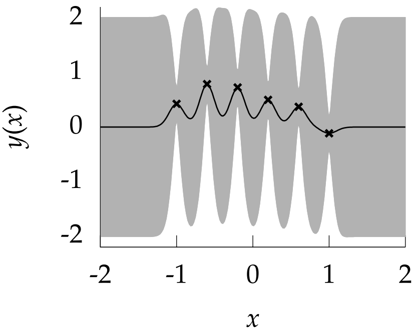
Can we determine length scales and noise levels from the data?

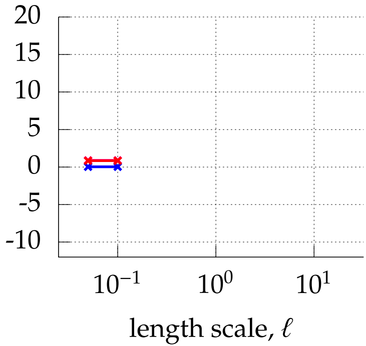


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

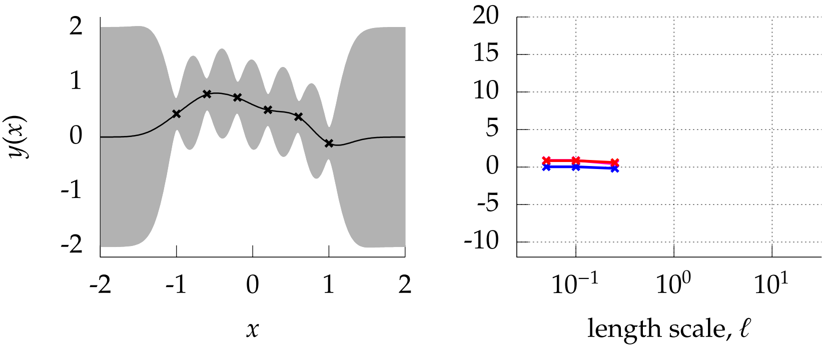




$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

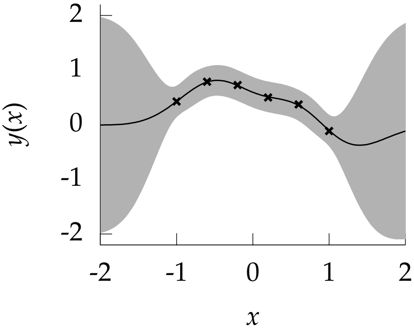
Can we determine length scales and noise levels from the data?

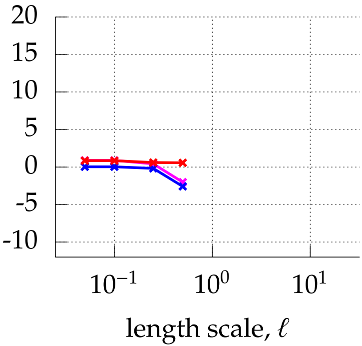


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

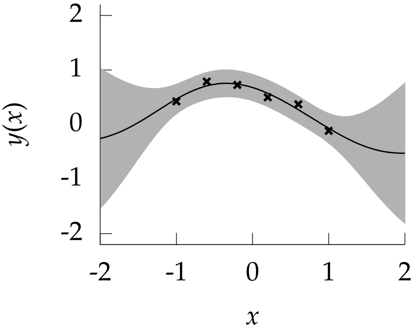


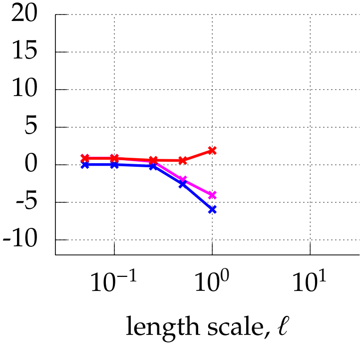


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

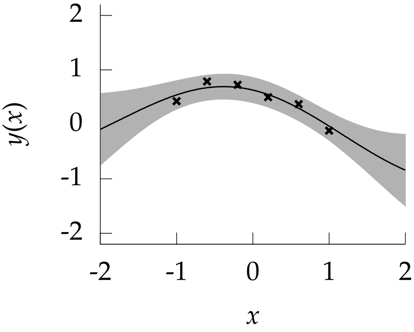


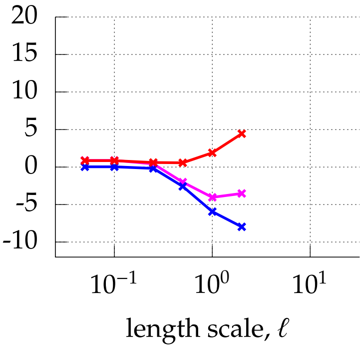


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

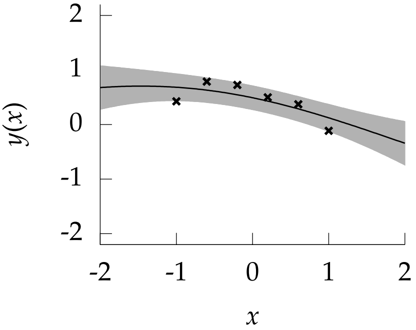


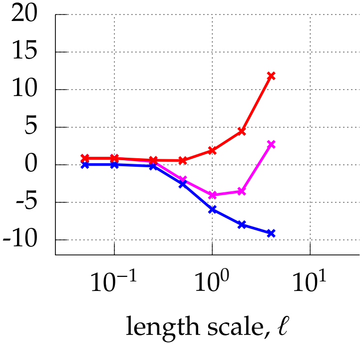


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

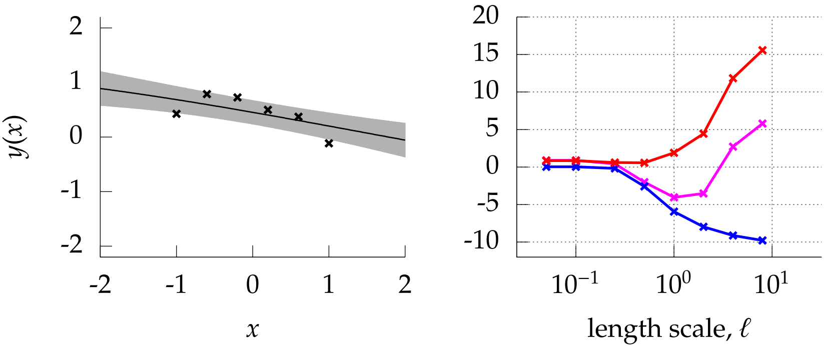




$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

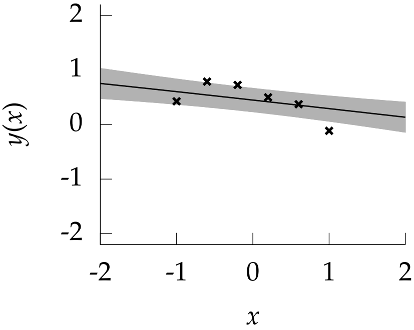
Can we determine length scales and noise levels from the data?

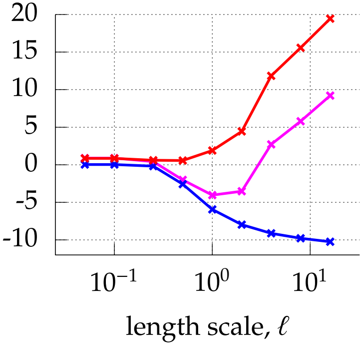


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

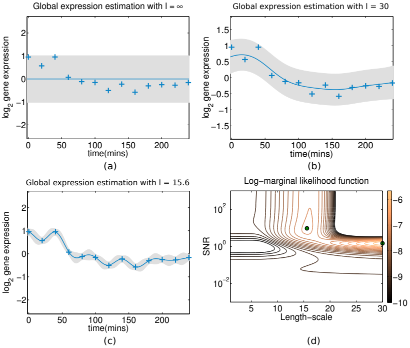
Can we determine length scales and noise levels from the data?



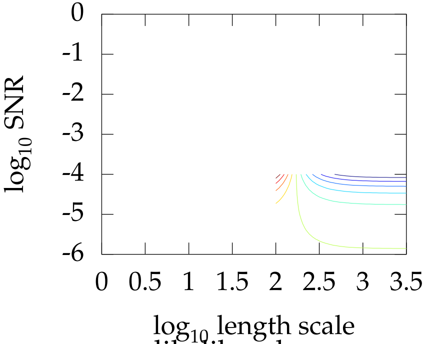


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top\mathbf{K}^{-1}\mathbf{y}}{2}$$

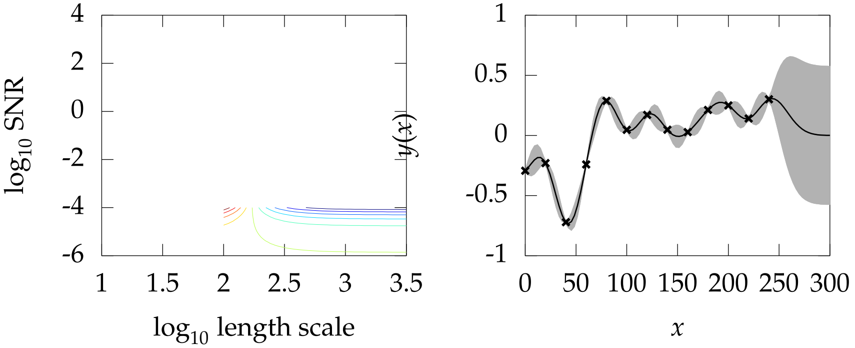
Gene Expression Example



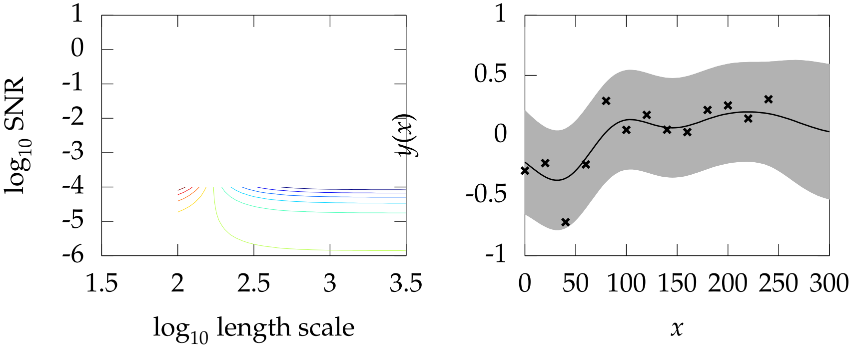
Data from Della Gatta et al. (2008). Application from Kalaitzis and Lawrence (2011).



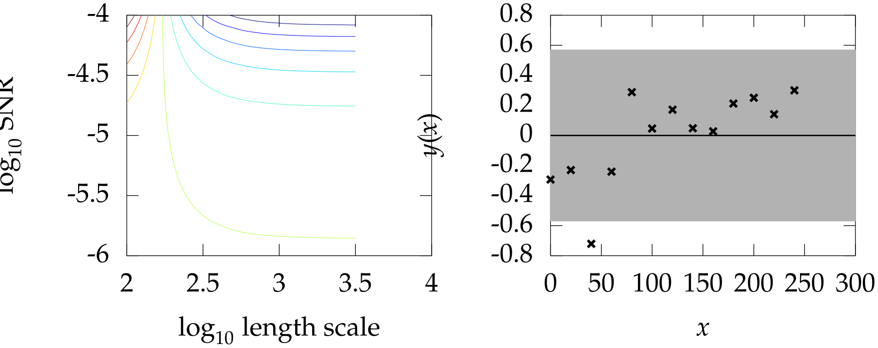
Contour plot of Gaussian process likelihood.



Optima: length scale of 1.2221 and $\log_{10}$ SNR of 1.9654 log likelihood is -0.22317.

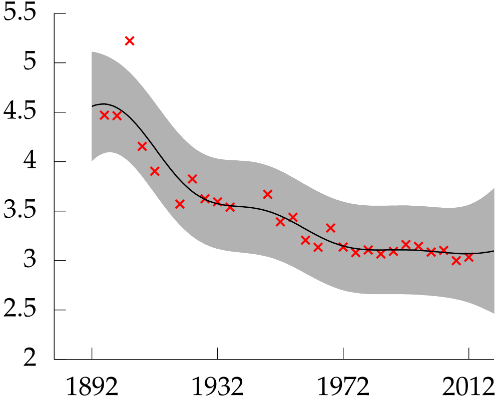


Optima: length scale of 1.5162 and $\log_{10}$ SNR of 0.21306 log likelihood is -0.23604.



Optima: length scale of 2.9886 and $\log_{10}$ SNR of -4.506 log likelihood is -2.1056.

Gaussian Process Fit to Olympic Marathon Data



Selecting Number and Location of Basis

- Need to choose
  1. location of centers

Selecting Number and Location of Basis

- Need to choose
  1. location of centers
  2. number of basis functions

Selecting Number and Location of Basis

- Need to choose
  1. location of centers
  2. number of basis functions
- Consider uniform spacing over a region:

$$k\left(x_i, x_j\right) = \gamma\Delta \sum_{k=1}^{m} \exp\left(-\frac{x_i^2 + x_j^2 - 2\mu_k\left(x_i + x_j\right) + 2\mu_k^2}{2\ell^2}\right),$$

Restrict analysis to 1-D input, x.

Uniform Basis Functions

- Set each center location to

$$\mu_k = a + \Delta\mu \cdot (k - 1).$$

Uniform Basis Functions

- Set each center location to

$$\mu_k = a + \Delta\mu \cdot (k - 1).$$

- Specify the basis functions in terms of their indices,

$$k\left(x_i, x_j\right) = \gamma \Delta\mu \sum_{k=0}^{m-1} \exp\left(- \frac{x_i^2 + x_j^2}{2\ell^2} - \frac{2\left(a + \Delta\mu \cdot k\right)\left(x_i + x_j\right) + 2\left(a + \Delta\mu \cdot k\right)^2}{2\ell^2} \right).$$

Infinite Basis Functions

- Take $\mu_0 = a$ and $\mu_m = b$ so $b = a + \Delta\mu \cdot (m - 1)$.

Infinite Basis Functions

- Take $\mu_0 = a$ and $\mu_m = b$ so $b = a + \Delta\mu \cdot (m - 1)$.
- Take limit as $\Delta\mu \to 0$ so $m \to \infty$

Infinite Basis Functions

- Take $\mu_0 = a$ and $\mu_m = b$ so $b = a + \Delta\mu \cdot (m - 1)$.
- Take limit as $\Delta\mu \to 0$ so $m \to \infty$

$$
\begin{aligned}
k(x_i, x_j) =& \gamma \int_a^b \exp\Bigg(- \frac{x_i^2 + x_j^2}{2\ell^2} \\
&+ \frac{2\left(\mu - \frac{1}{2}\left(x_i + x_j\right)\right)^2 - \frac{1}{2}\left(x_i + x_j\right)^2}{2\ell^2}\Bigg) \mathrm{d}\mu,
\end{aligned}
$$

where we have used $k \cdot \Delta\mu \to \mu$.

Result

- Performing the integration leads to

$$k(x_i,x_j) = \gamma \frac{\sqrt{\pi \ell^2}}{2} \exp\left(- \frac{\left(x_i - x_j\right)^2}{4\ell^2} \right)$$
$$\times \left[\operatorname{erf}\left(\frac{\left(b - \frac{1}{2}\left(x_i + x_j\right)\right)}{\ell} \right) - \operatorname{erf}\left(\frac{\left(a - \frac{1}{2}\left(x_i + x_j\right)\right)}{\ell} \right) \right],$$

Result

- Performing the integration leads to

$$k(x_i,x_j) = \gamma \frac{\sqrt{\pi\ell^2}}{2} \exp\left(-\frac{\left(x_i - x_j\right)^2}{4\ell^2}\right)$$
$$\times \left[\operatorname{erf}\left(\frac{\left(b - \frac{1}{2}\left(x_i + x_j\right)\right)}{\ell}\right) - \operatorname{erf}\left(\frac{\left(a - \frac{1}{2}\left(x_i + x_j\right)\right)}{\ell}\right)\right],$$

- Now take limit as $a \to -\infty$ and $b \to \infty$

Result

- Performing the integration leads to

$$k(x_i,x_j) = \gamma \frac{\sqrt{\pi\ell^2}}{2} \exp\left(-\frac{\left(x_i - x_j\right)^2}{4\ell^2}\right)$$
$$\times \left[\text{erf}\left(\frac{\left(b - \frac{1}{2}\left(x_i + x_j\right)\right)}{\ell}\right) - \text{erf}\left(\frac{\left(a - \frac{1}{2}\left(x_i + x_j\right)\right)}{\ell}\right)\right],$$

- Now take limit as $a \to -\infty$ and $b \to \infty$

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left(x_i - x_j\right)^2}{4\ell^2}\right).$$

where $\alpha = \gamma \sqrt{\pi\ell^2}$.

Infinite Feature Space

- An RBF model with infinite basis functions is a Gaussian process.

Infinite Feature Space

- An RBF model with infinite basis functions is a Gaussian process.
- The covariance function is given by the exponentiated quadratic covariance function.

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left(x_i - x_j\right)^2}{4\ell^2}\right).$$

where $\alpha = \gamma\sqrt{\pi\ell^2}$.

Infinite Feature Space

- An RBF model with infinite basis functions is a Gaussian process.
- The covariance function is the exponentiated quadratic.
- **Note:** The functional form for the covariance function and basis functions are similar.
  - this is a special case,
  - in general they are very different

Similar results can obtained for multi-dimensional input models Williams (1998); Neal (1996).

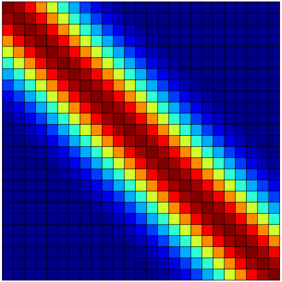
Covariance Functions

Where did this covariance matrix come from?

Exponentiated Quadratic Kernel Function (RBF, Squared Exponential, Gaussian)

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function **x**.
- For the example above it was based on Euclidean distance.
- The covariance function is also know as a kernel.



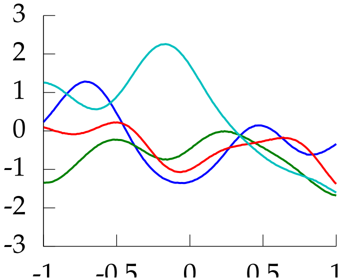
Covariance Functions

Where did this covariance matrix come from?

Exponentiated Quadratic Kernel Function (RBF, Squared Exponential, Gaussian)

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function $\mathbf{x}$.
- For the example above it was based on Euclidean distance.
- The covariance function is also know as a kernel.



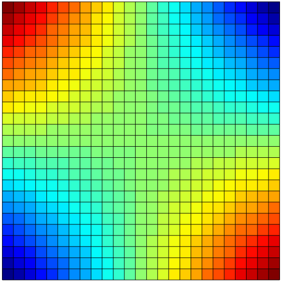
Covariance Functions

Linear Covariance Function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \mathbf{x}^\top \mathbf{x}'$$

- Bayesian linear regression.

$$\alpha = 1$$



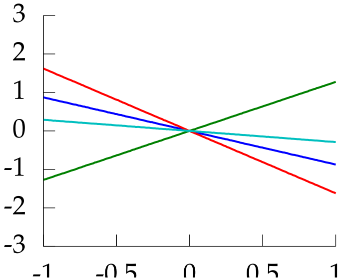
Covariance Functions

Linear Covariance Function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \mathbf{x}^\top \mathbf{x}'$$

- Bayesian linear regression.

$$\alpha = 1$$



Covariance Functions

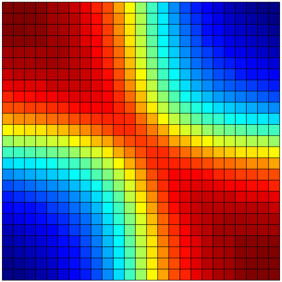
MLP Covariance Function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \operatorname{asin}\left(\frac{w\mathbf{x}^\top \mathbf{x}' + b}{\sqrt{w\mathbf{x}^\top \mathbf{x} + b + 1}\sqrt{w\mathbf{x}'^\top \mathbf{x}' + b + 1}}\right)$$

- Based on infinite neural network model.

$$w = 40$$

$$b = 4$$



Covariance Functions

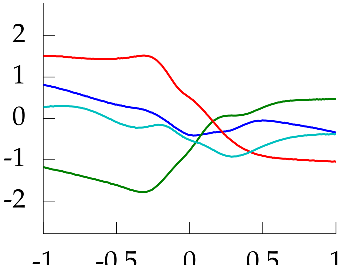
MLP Covariance Function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \operatorname{asin}\left(\frac{w\mathbf{x}^\top \mathbf{x}' + b}{\sqrt{w\mathbf{x}^\top \mathbf{x} + b + 1}\sqrt{w\mathbf{x}'^\top \mathbf{x}' + b + 1}}\right)$$

- Based on infinite neural network model.

$$w = 40$$

$$b = 4$$



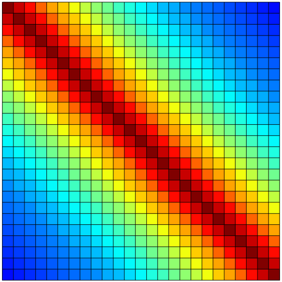
Covariance Functions

Where did this covariance matrix come from?

Ornstein-Uhlenbeck (stationary Gauss-Markov) covariance function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function $\mathbf{x}$.



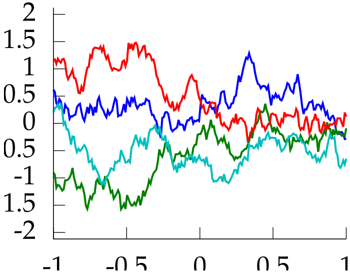
Covariance Functions

Where did this covariance matrix come from?

Ornstein-Uhlenbeck (stationary Gauss-Markov) covariance function

$$k\left(\mathbf{x}, \mathbf{x}'\right) = \alpha \exp\left(-\frac{\left|\mathbf{x} - \mathbf{x}'\right|}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function $\mathbf{x}$.



References I

C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006. [Google Books] .

G. Della Gatta, M. Bansal, A. Ambesi-Impiombato, D. Antonini, C. Missero, and D. di Bernardo. Direct targets of the trp63 transcription factor revealed by a combination of gene expression profiling and reverse engineering. *Genome Research*, 18(6):939–948, Jun 2008. [URL]. [DOI].

A. A. Kalaitzis and N. D. Lawrence. A simple approach to ranking differentially expressed gene expression time courses through Gaussian process regression. *BMC Bioinformatics*, 12(180), 2011. [DOI].

R. M. Neal. *Bayesian Learning for Neural Networks*. Springer, 1996. Lecture Notes in Statistics 118.

J. Oakley and A. O'Hagan. Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89(4):769–784, 2002.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2006. [Google Books] .

C. K. I. Williams. Computation with infinite neural networks. *Neural Computation*, 10(5):1203–1216, 1998.