GAUSSIAN PROCESS FOR REGRESSION

Hsin-Min Lu
盧信銘
台大資管系

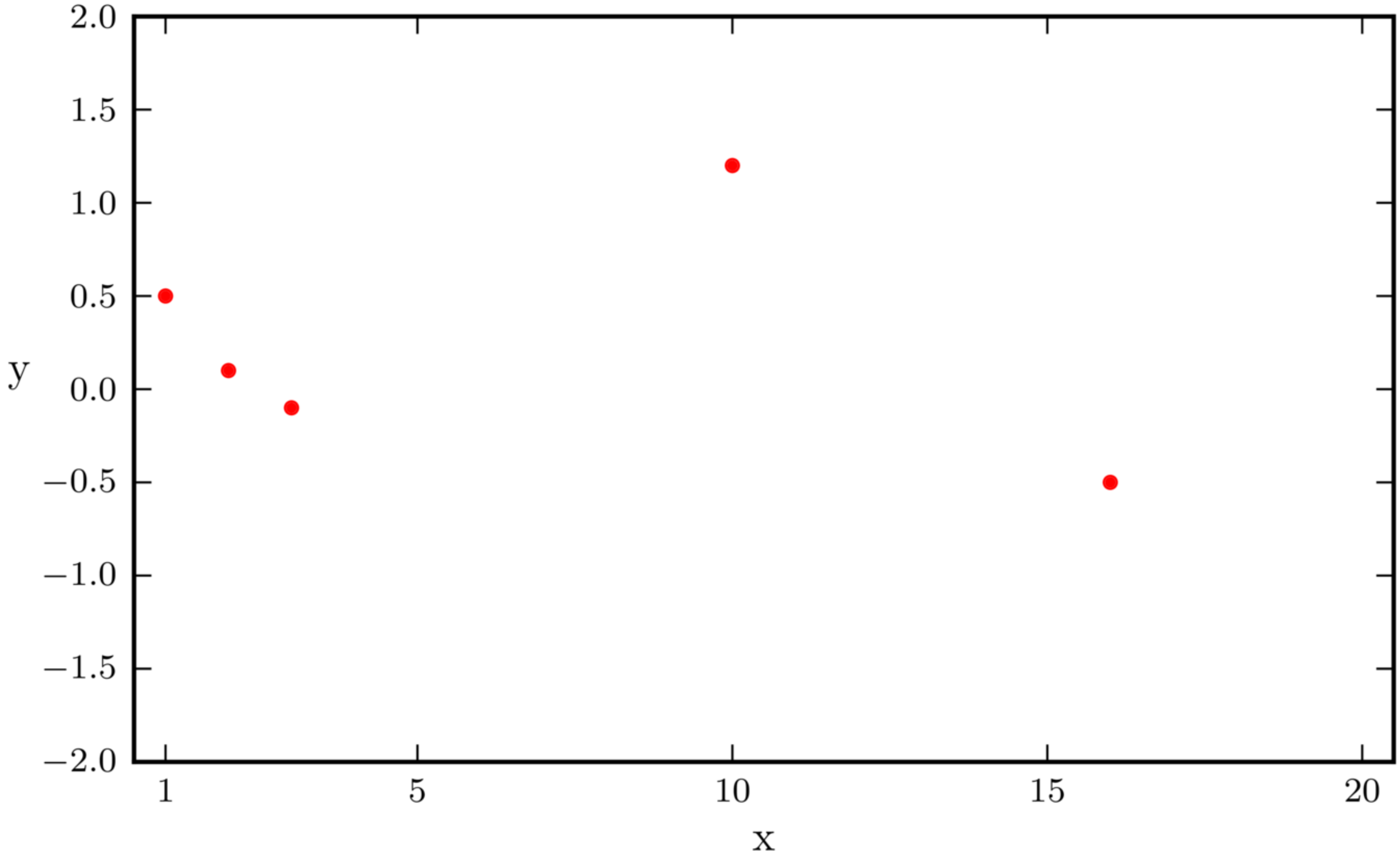
Majority of slides were borrowed from Dr. Richard E. Turner

Gaussian Processes

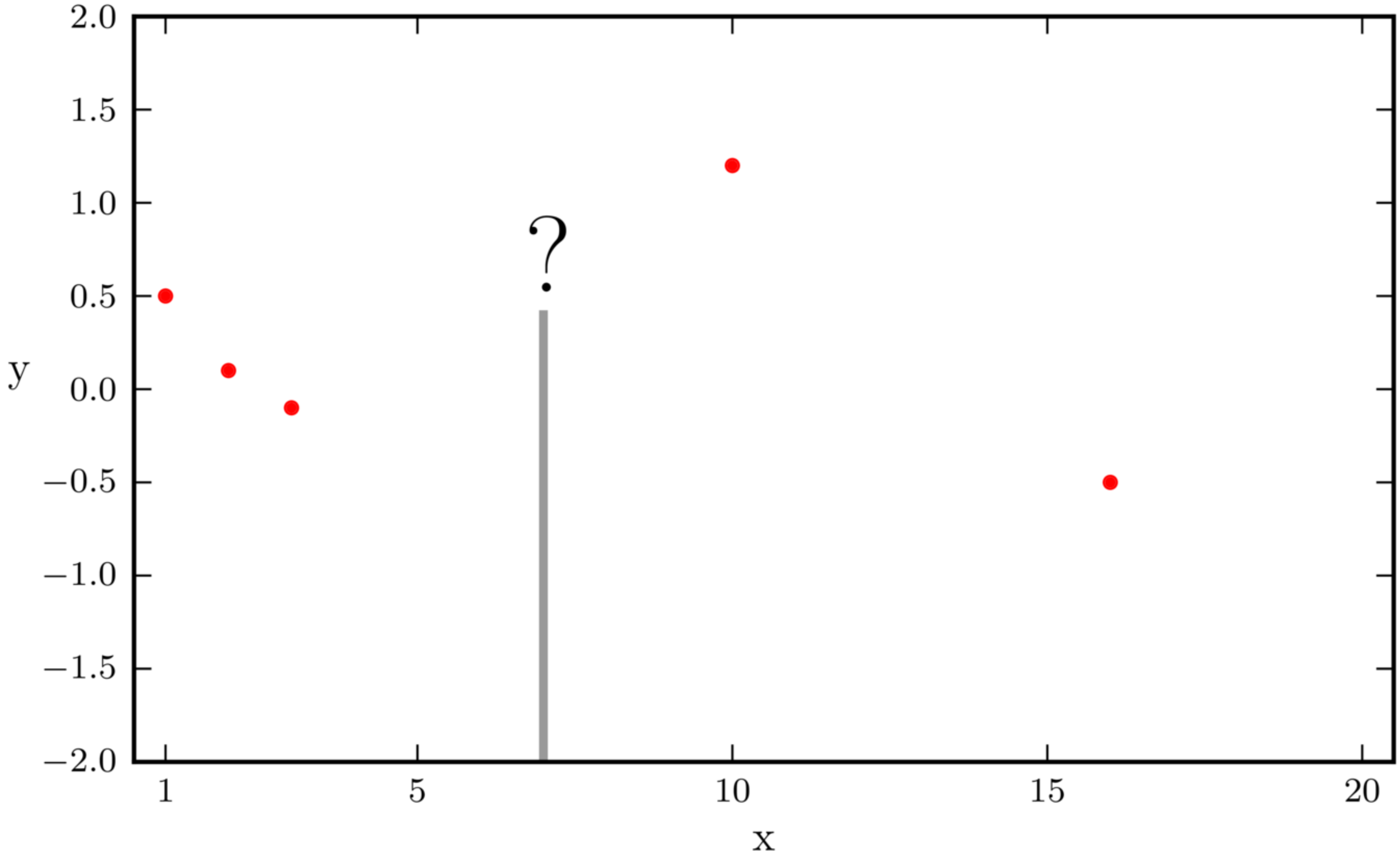
- Gaussian processes (GPs) provide a principled, practical, probabilistic approach to learning in kernel machines.
- The “kernel” plays the role of covariance matrix in GPs.
- Can be used to learn nonlinear regression and classification problems.
- GPs are nonparametric models!
- We will start with the nonlinear regression problems.

- Note: Gaussian Process Regression (GPR) is not Linear Regression or Ridge Regression. GPR is a nonlinear model. It is more close to support vector regressions than to linear regressions.

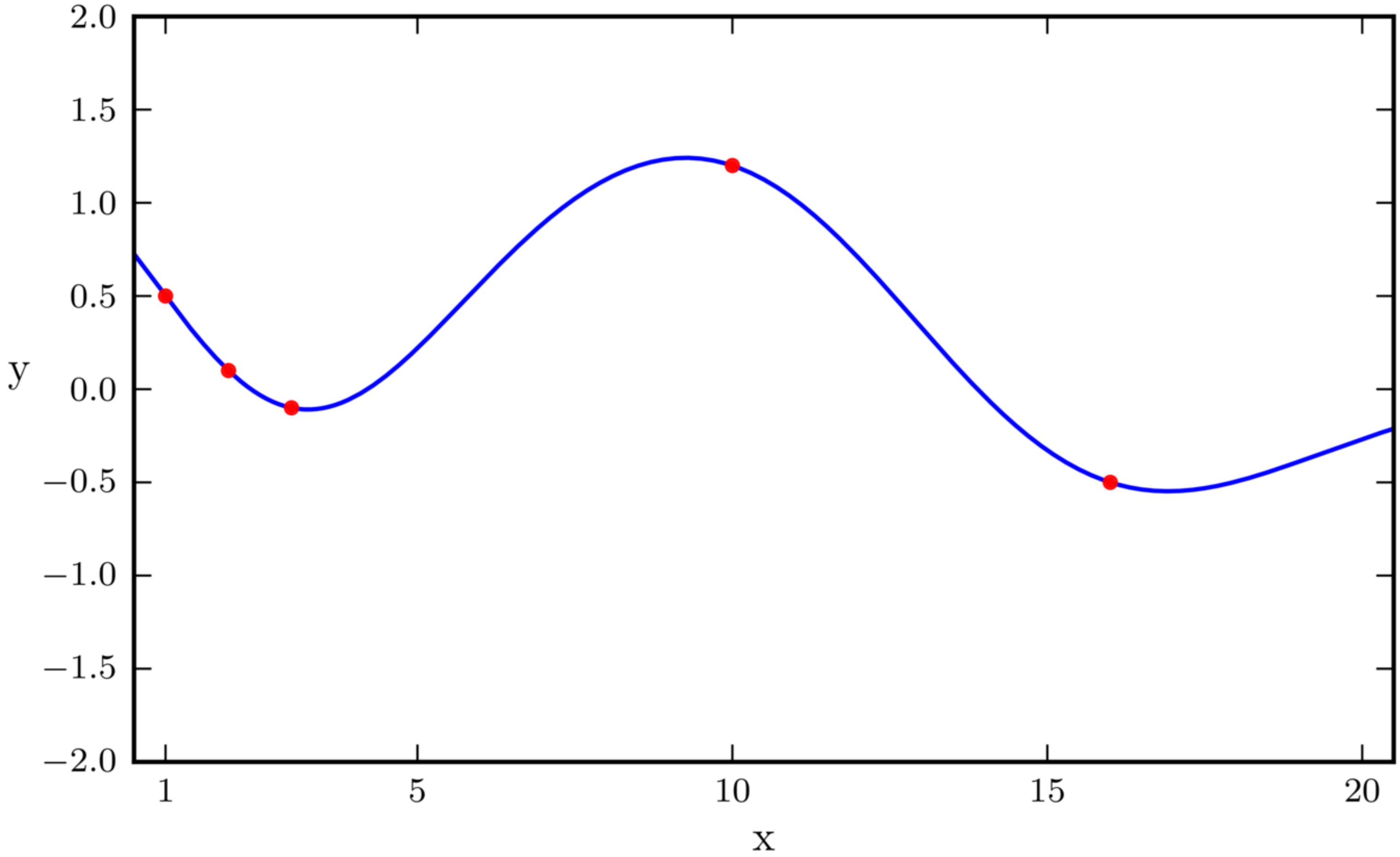
Nonlinear Regression Problem



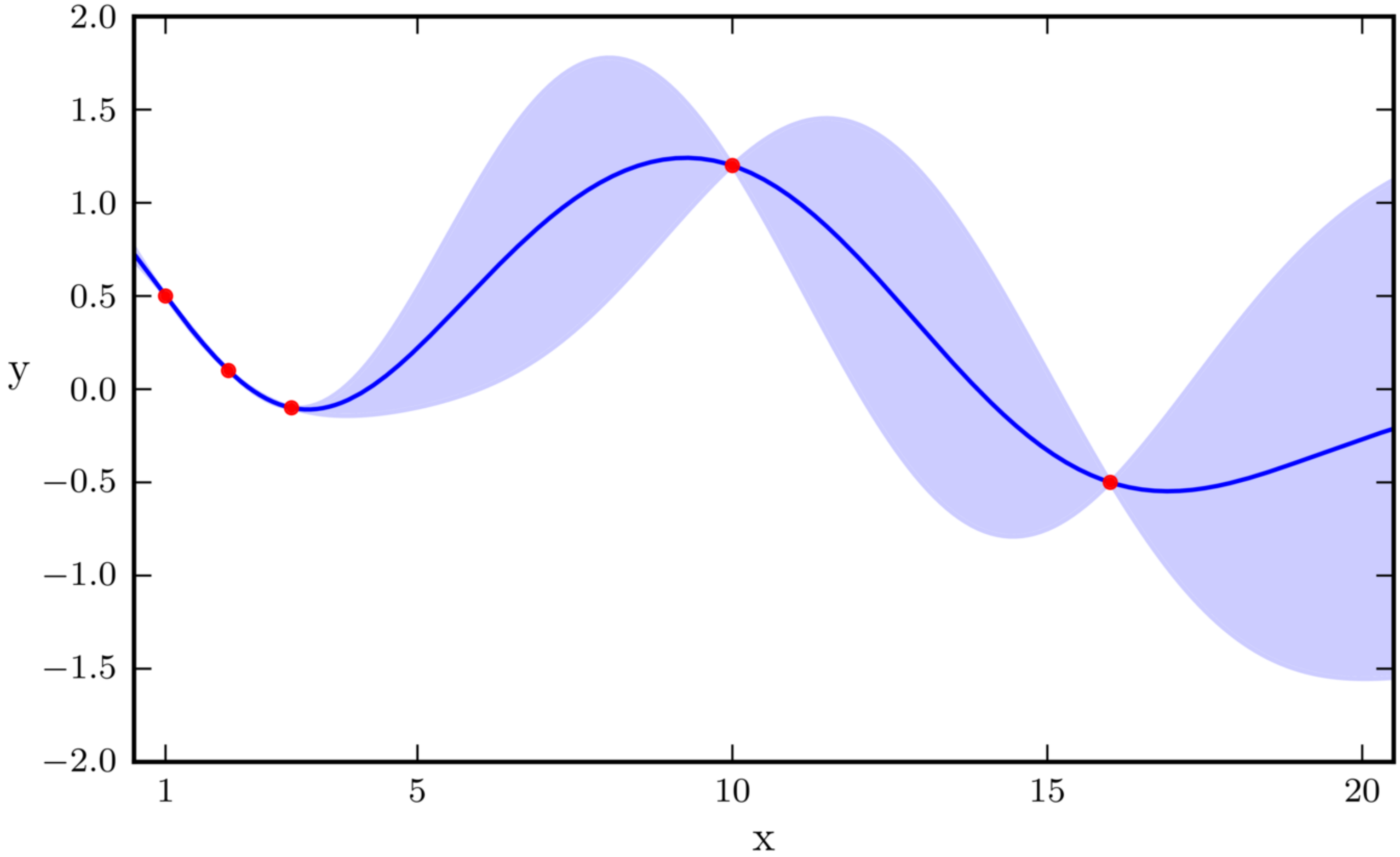
Nonlinear Regression Problem



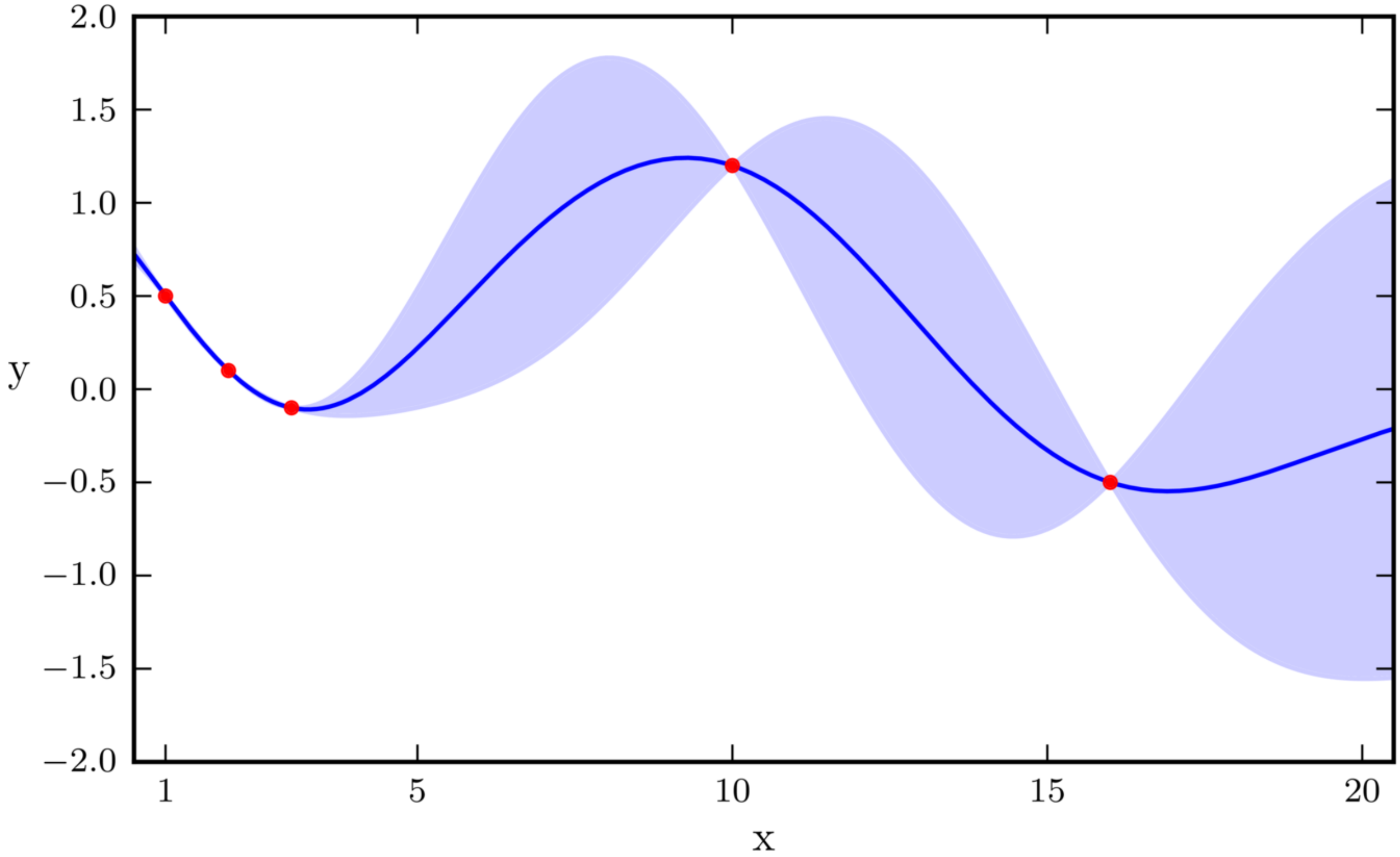
Nonlinear Regression Problem



Nonlinear Regression Problem



Nonlinear Regression Problem

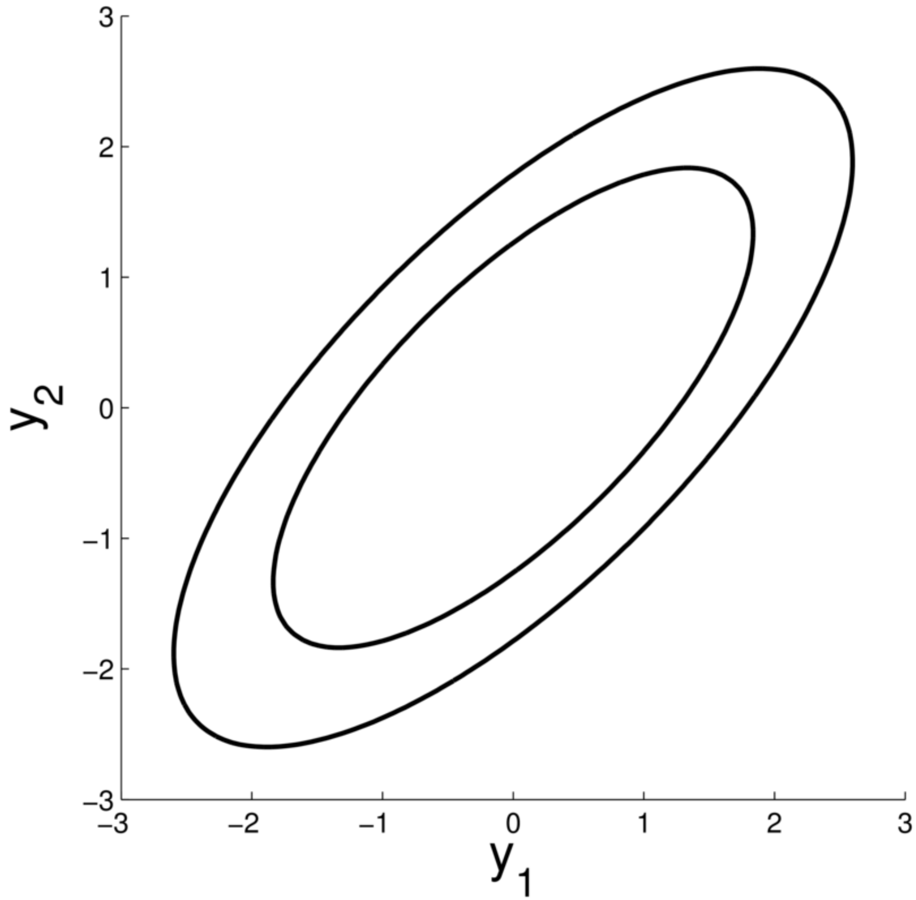


Can we do this with a plain old Gaussian?

Gaussian Distribution

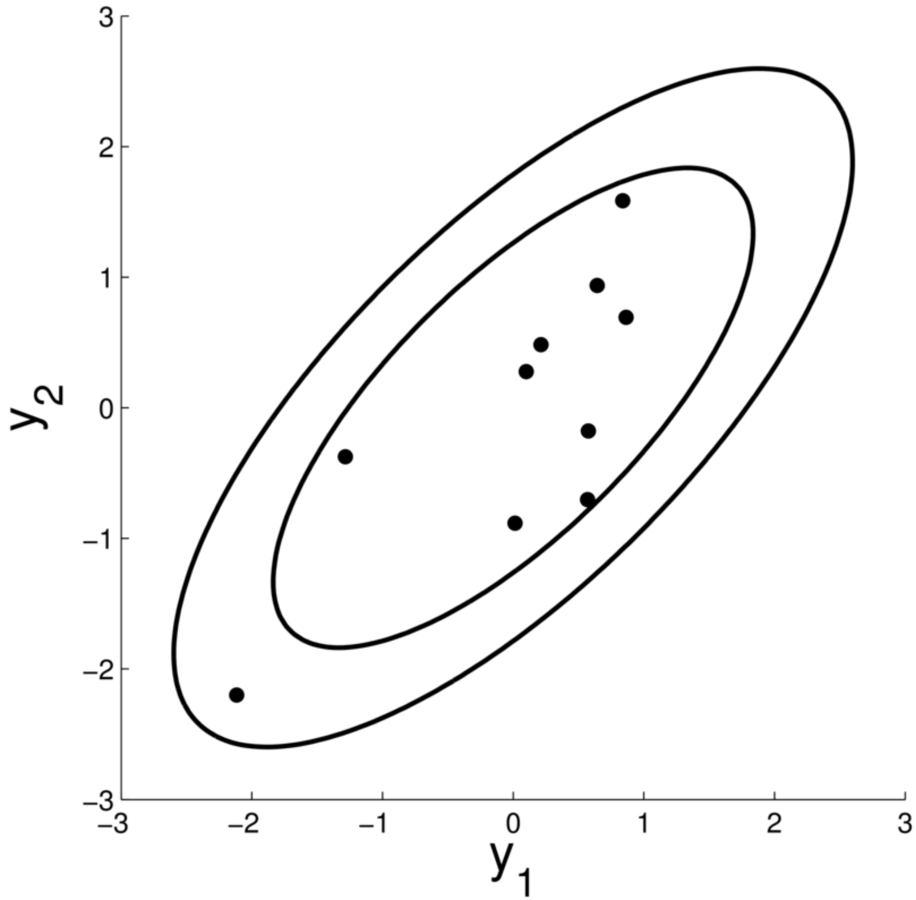
$$f_{\mathbf{X}}(x_1,\ldots,x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}$$

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$



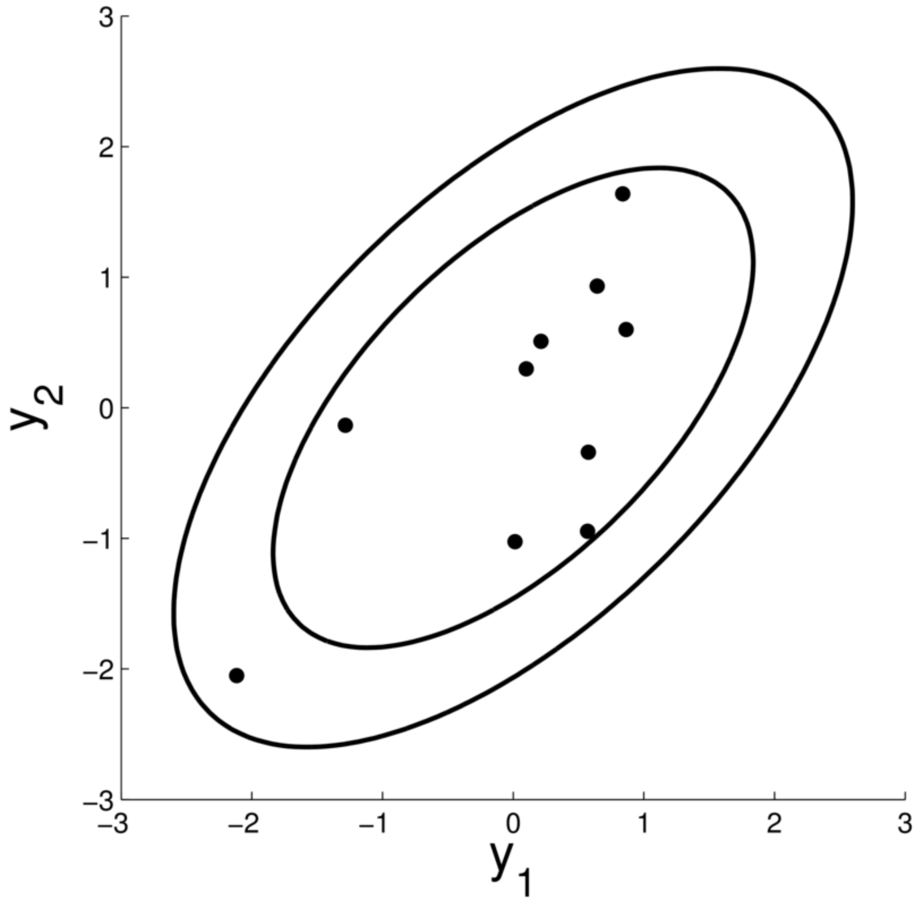
Gaussian Distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\top\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$



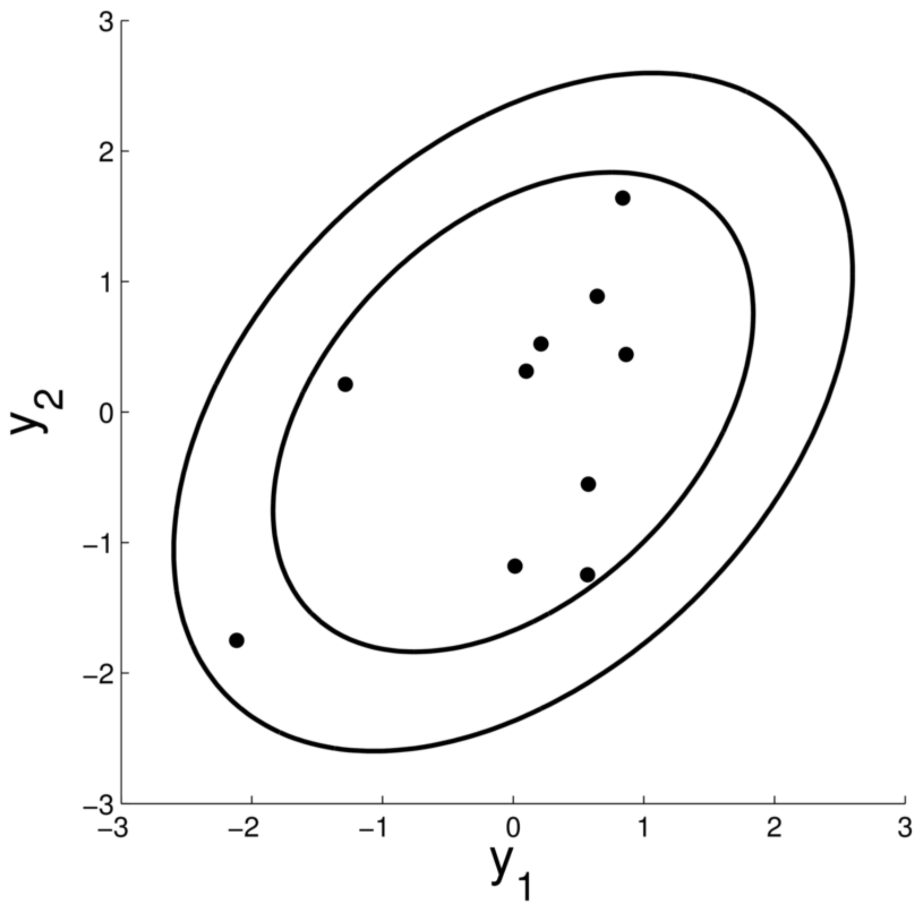
Gaussian Distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^{\top}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix}$$



Gaussian Distribution

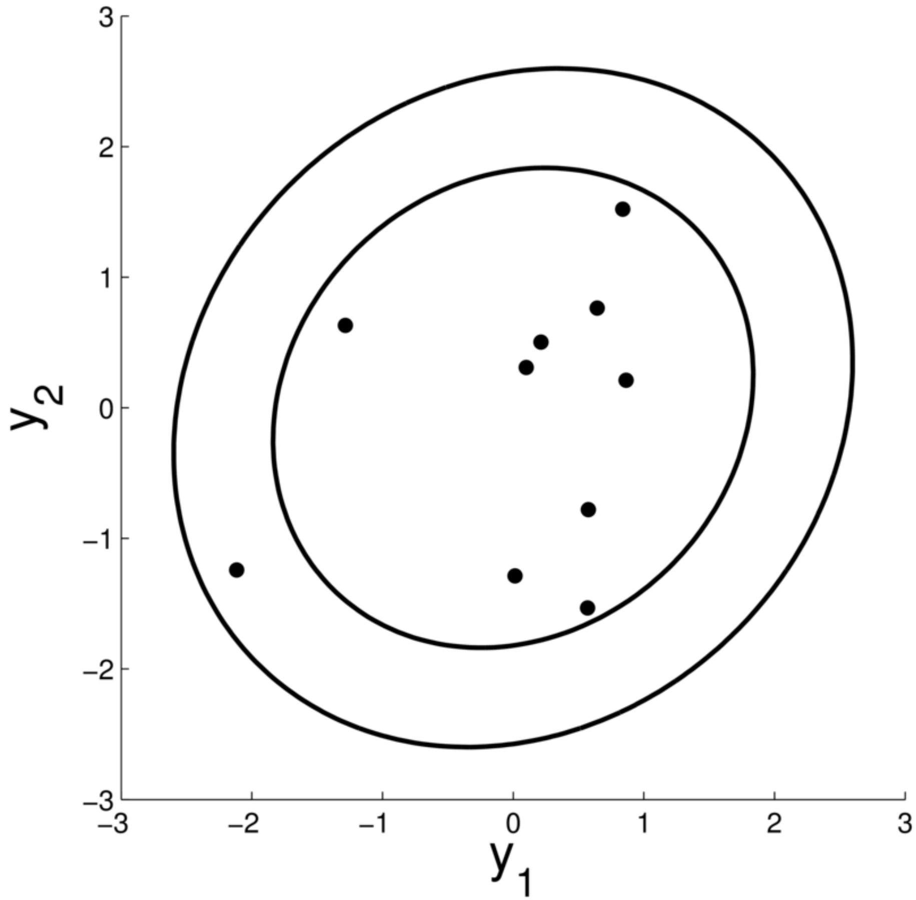
$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\top\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .4 \\ .4 & 1 \end{bmatrix}$$



Gaussian Distribution

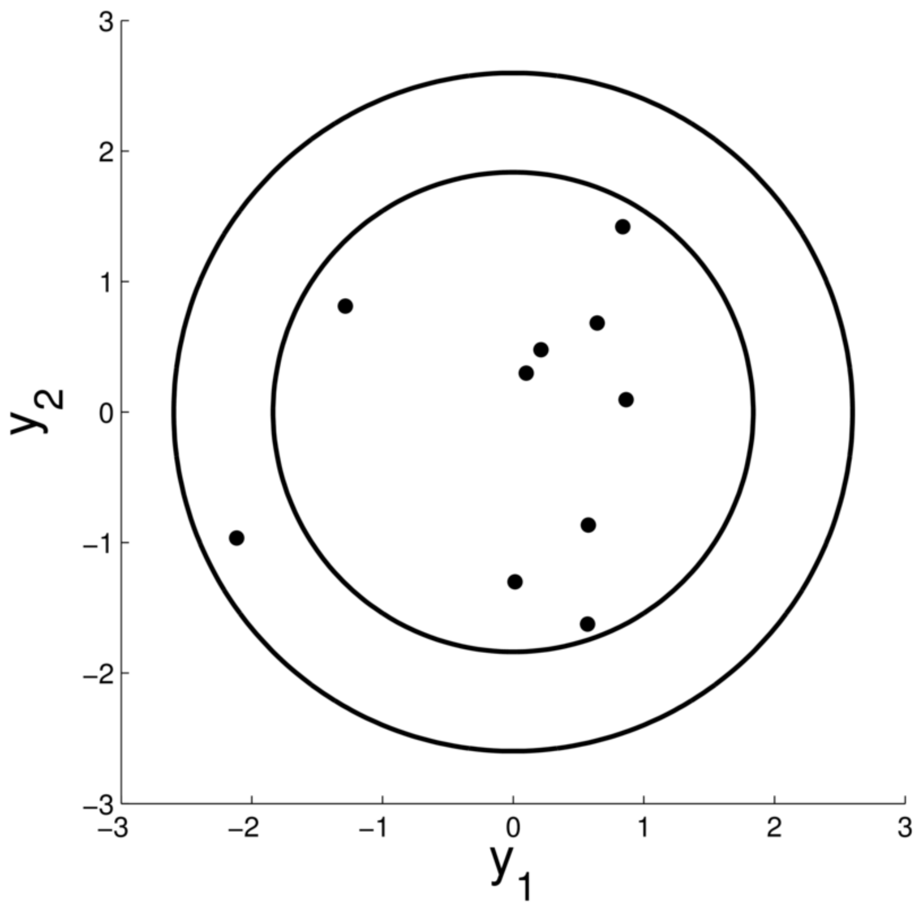
$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\top\Sigma^{-1}\mathbf{y}\right)$$

$$\Sigma = \begin{bmatrix} 1 & .1 \\ .1 & 1 \end{bmatrix}$$



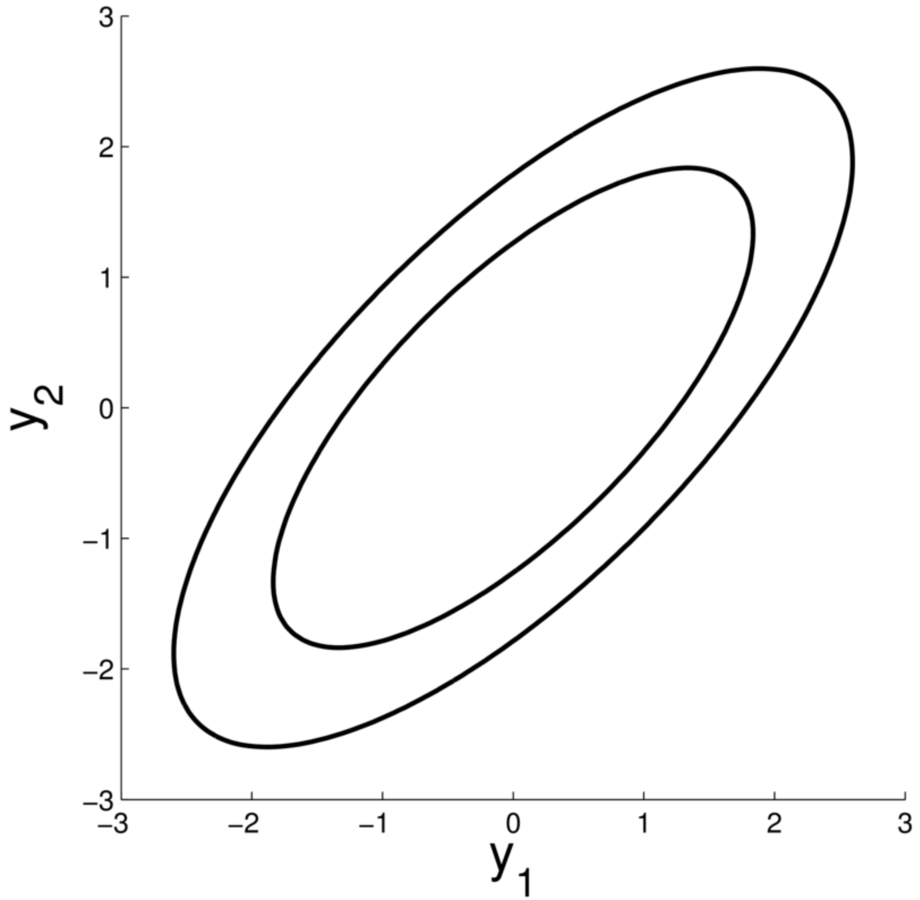
Gaussian Distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\top\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



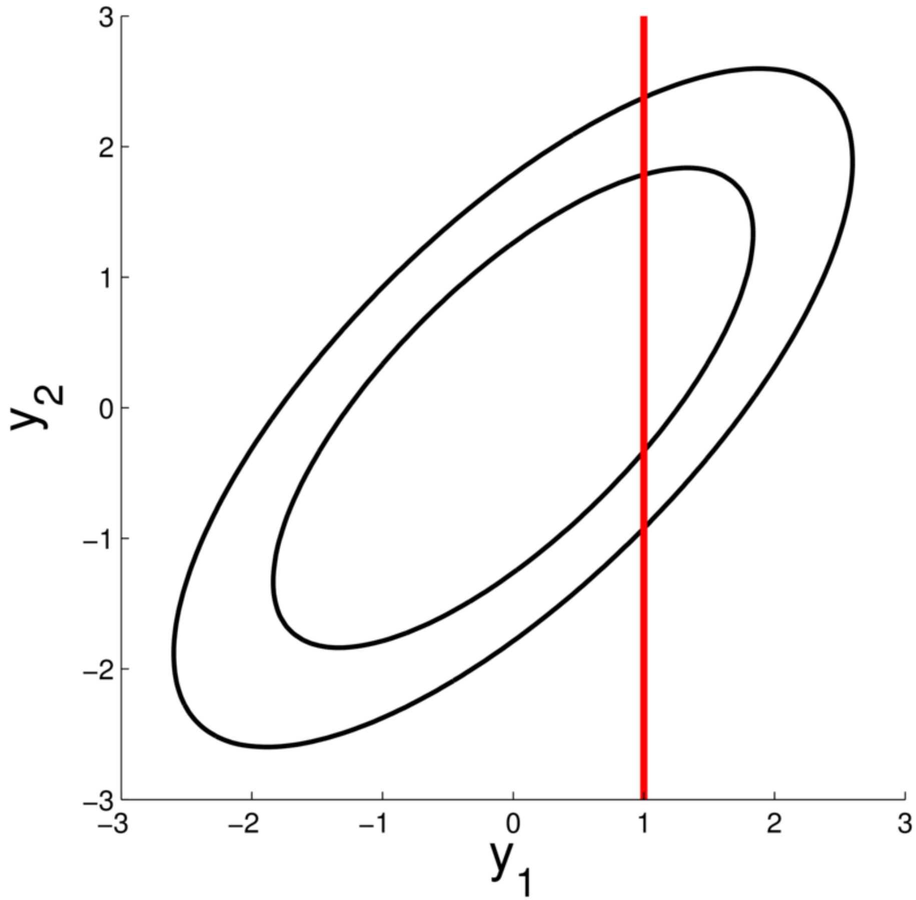
Gaussian Distribution: Conditional Distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\top\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$



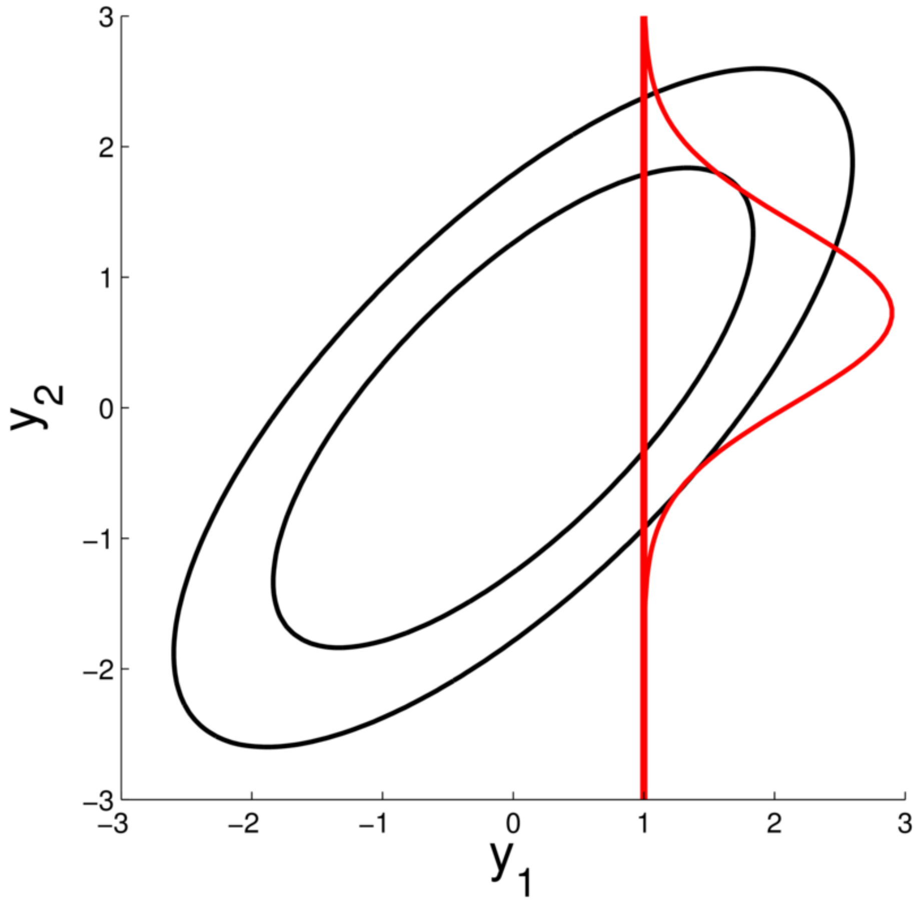
Gaussian Distribution: Conditional Distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$



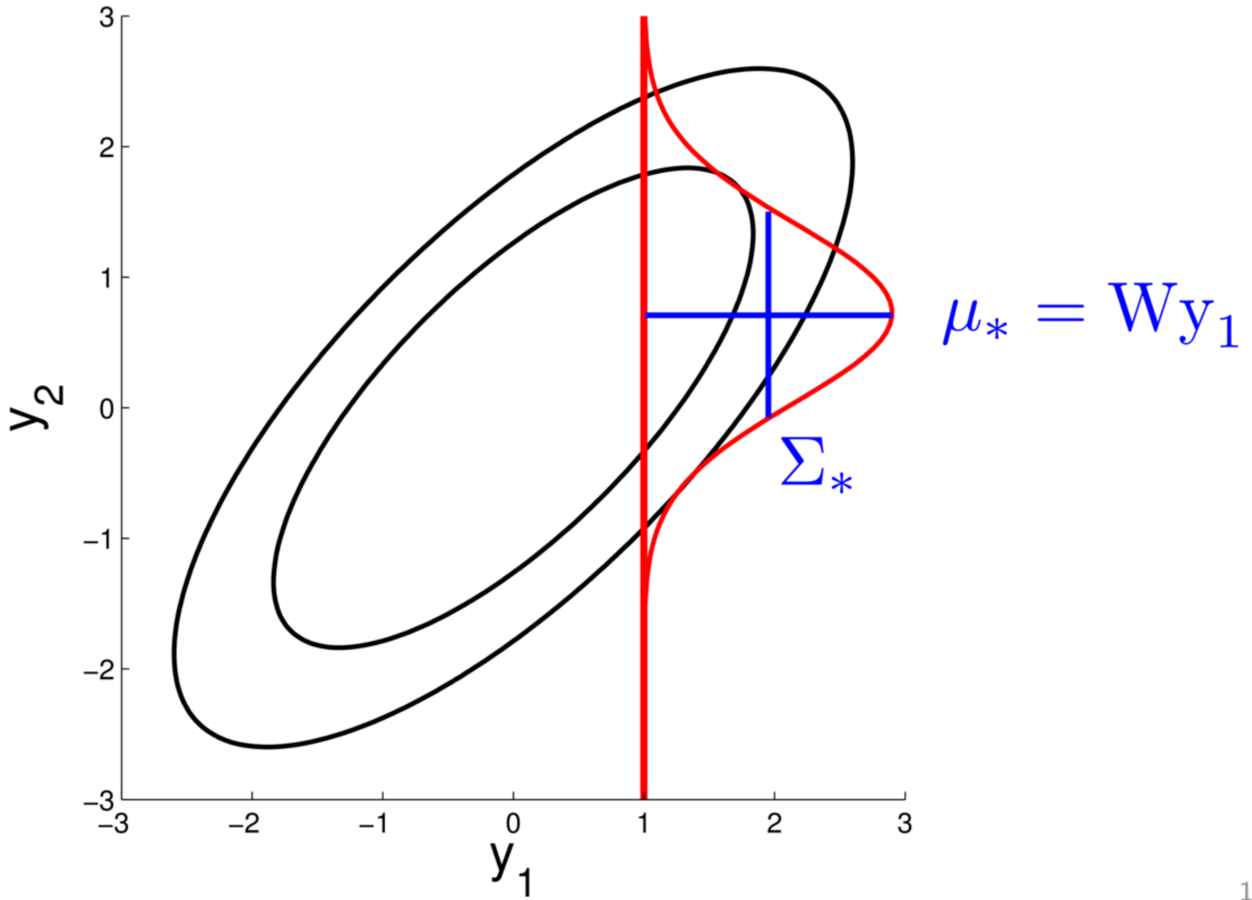
Gaussian Distribution: Conditional Distribution

$$p(\mathrm{y}_2|\mathrm{y}_1, \Sigma) \propto \exp\left(-\tfrac{1}{2}(\mathrm{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathrm{y}_2 - \mu_*)\right)$$



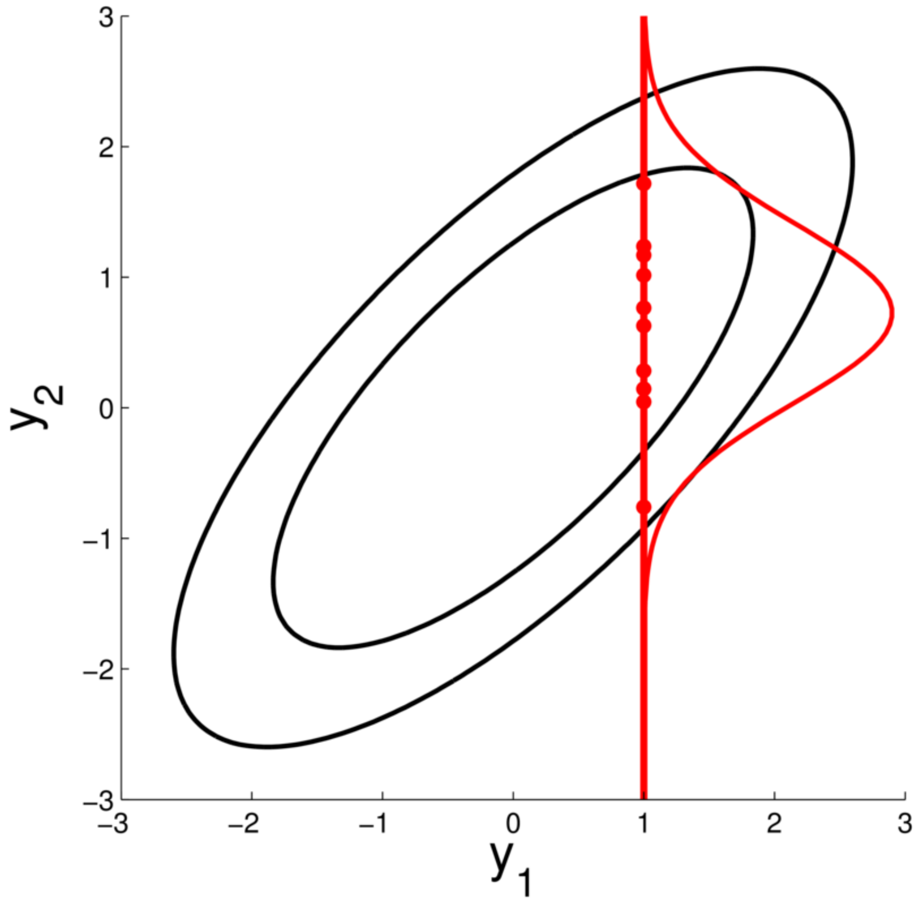
Gaussian Distribution: Conditional Distribution

$$p(\mathrm{y}_2|\mathrm{y}_1, \Sigma) \propto \exp\left(-\frac{1}{2}(\mathrm{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathrm{y}_2 - \mu_*)\right)$$



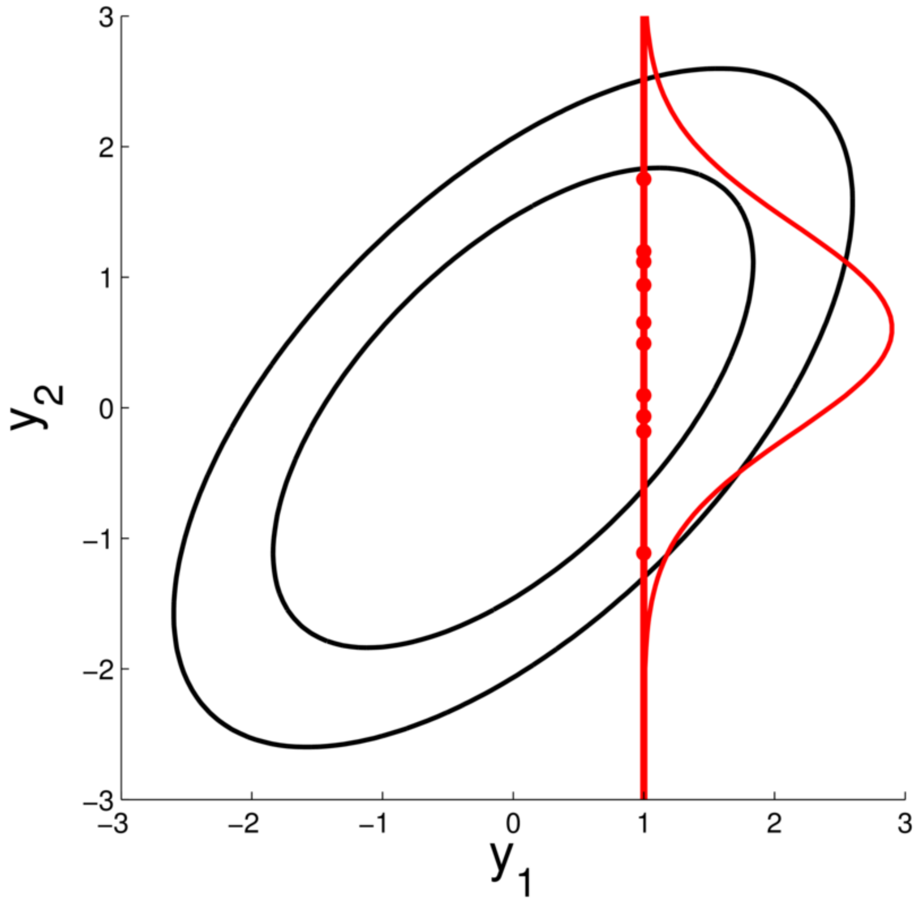
Gaussian Distribution

$$p(y_2|y_1, \Sigma) \propto \exp\left(-\frac{1}{2}(y_2 - \mu_*)\Sigma_*{}^{-1}(y_2 - \mu_*)\right)$$



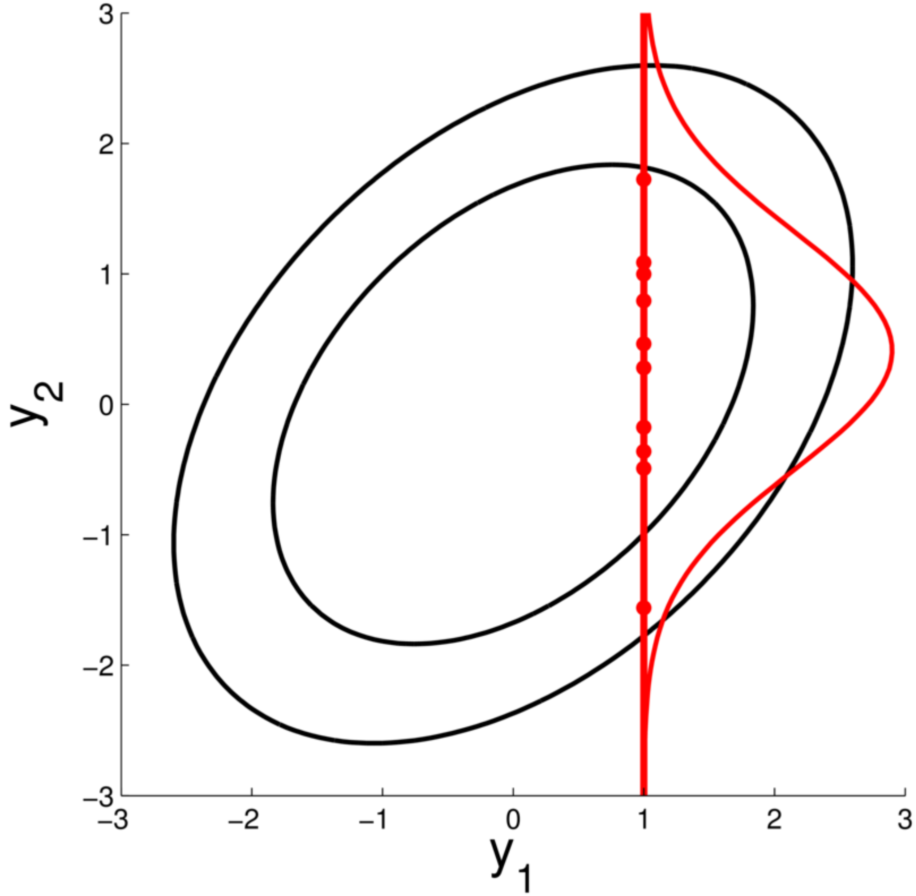
Gaussian Distribution: Conditional Distribution

$$p(\mathrm{y}_2|\mathrm{y}_1, \Sigma) \propto \exp\left(-\frac{1}{2}(\mathrm{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathrm{y}_2 - \mu_*)\right)$$



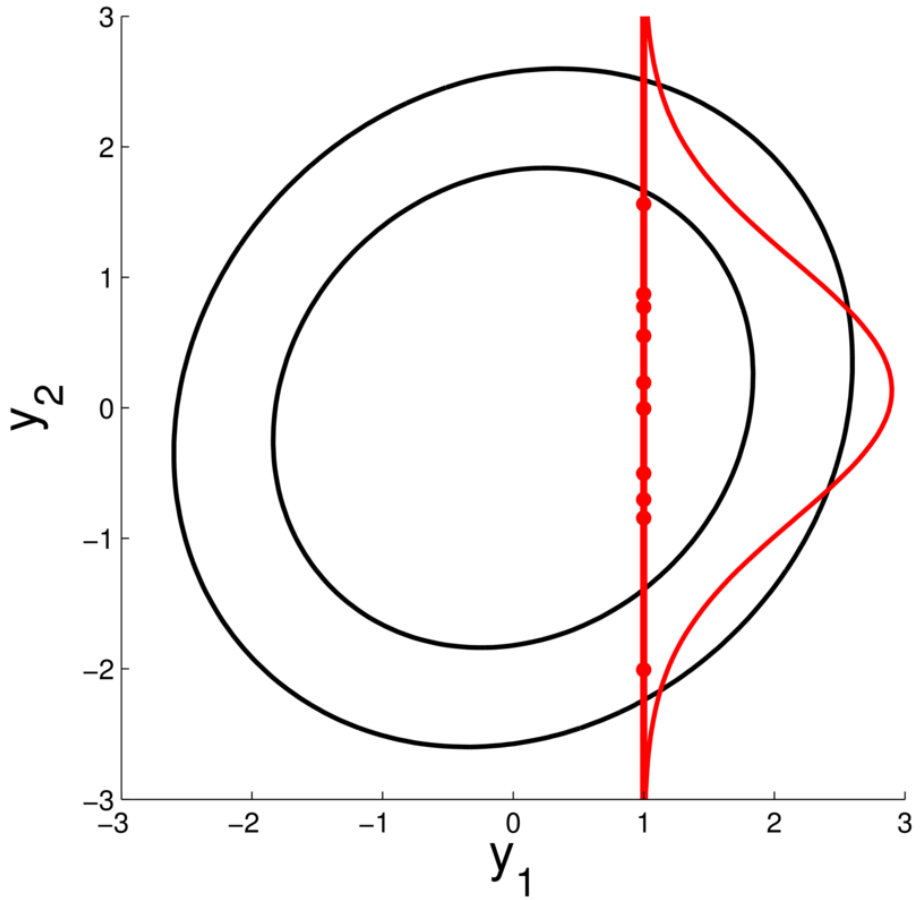
Gaussian Distribution: Conditional Distribution

$$p(\mathrm{y}_2|\mathrm{y}_1,\Sigma) \propto \exp\left(-\frac{1}{2}(\mathrm{y}_2-\mu_*)\Sigma_*{}^{-1}(\mathrm{y}_2-\mu_*)\right)$$



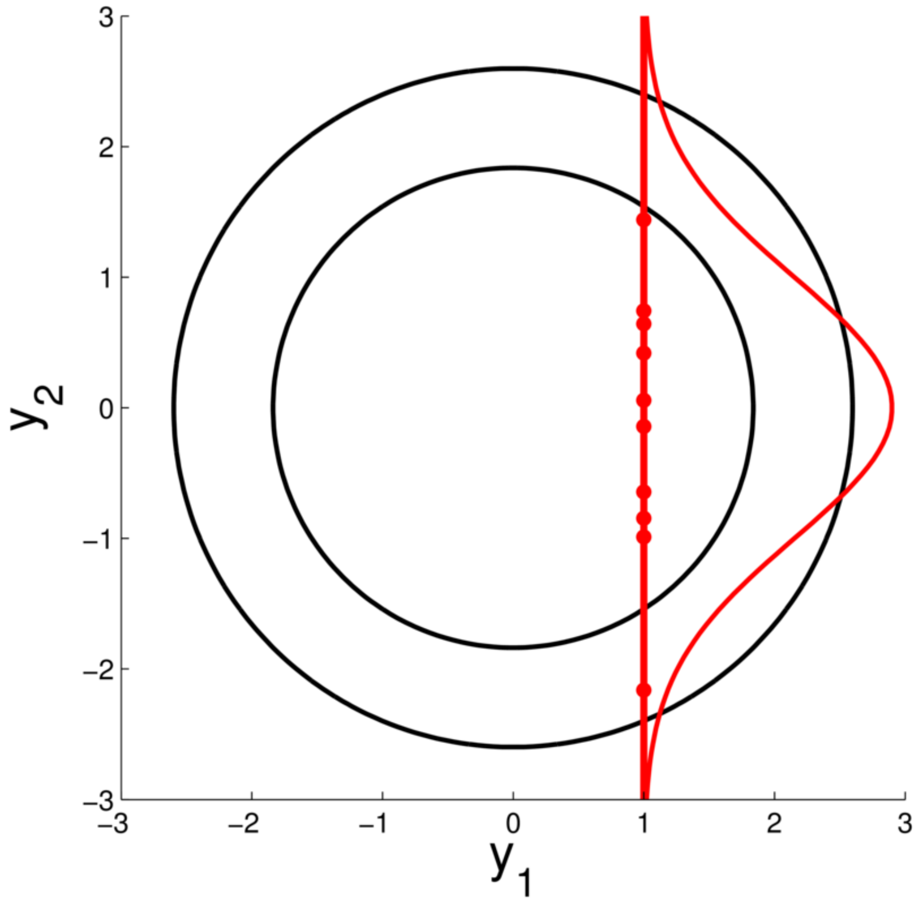
Gaussian Distribution: Conditional Distribution

$$p(\mathrm{y}_2|\mathrm{y}_1, \Sigma) \propto \exp\left(-\frac{1}{2}(\mathrm{y}_2 - \mu_*)\Sigma_*^{-1}(\mathrm{y}_2 - \mu_*)\right)$$

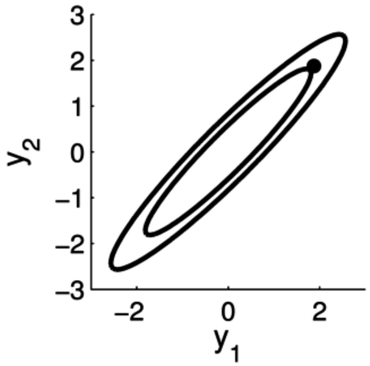


Gaussian Distribution: Conditional Distribution

$$p(\mathrm{y}_2|\mathrm{y}_1,\Sigma) \propto \exp\left(-\frac{1}{2}(\mathrm{y}_2-\mu_*)\Sigma_*{}^{-1}(\mathrm{y}_2-\mu_*)\right)$$

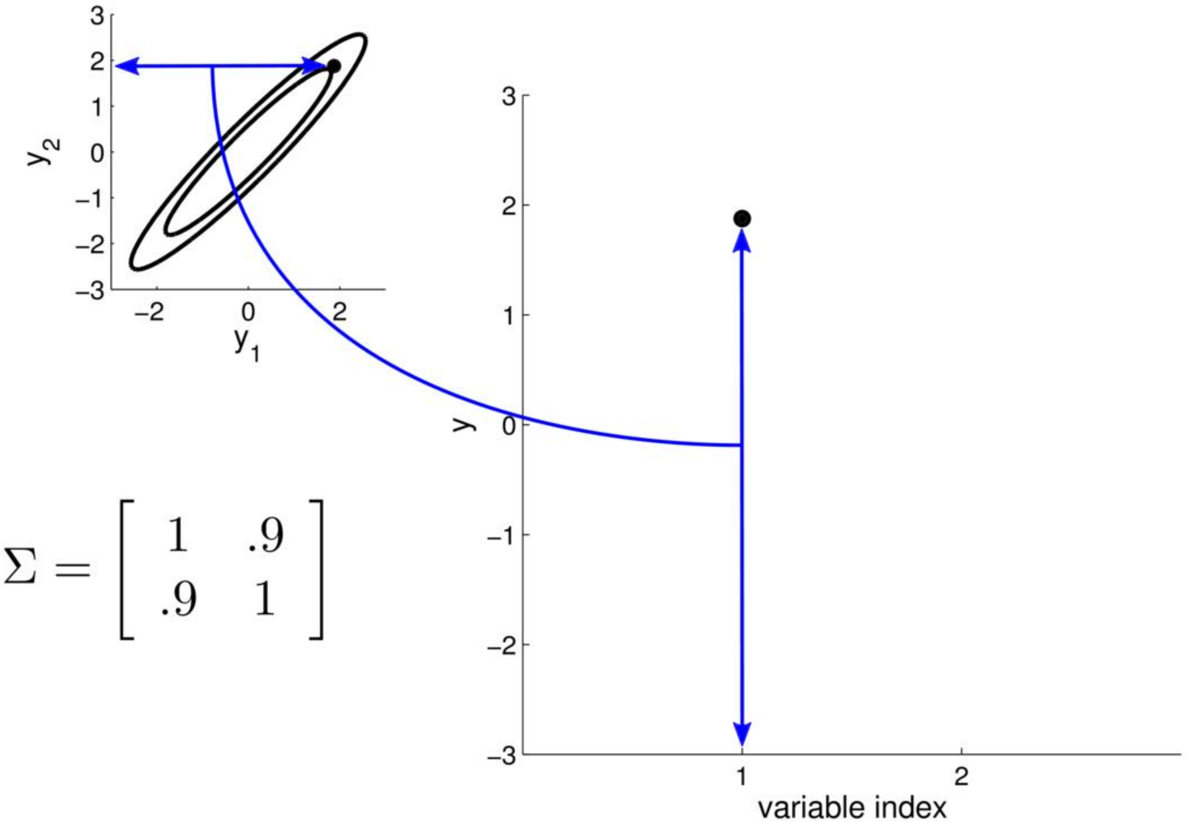


Visualizing Gaussian Processes



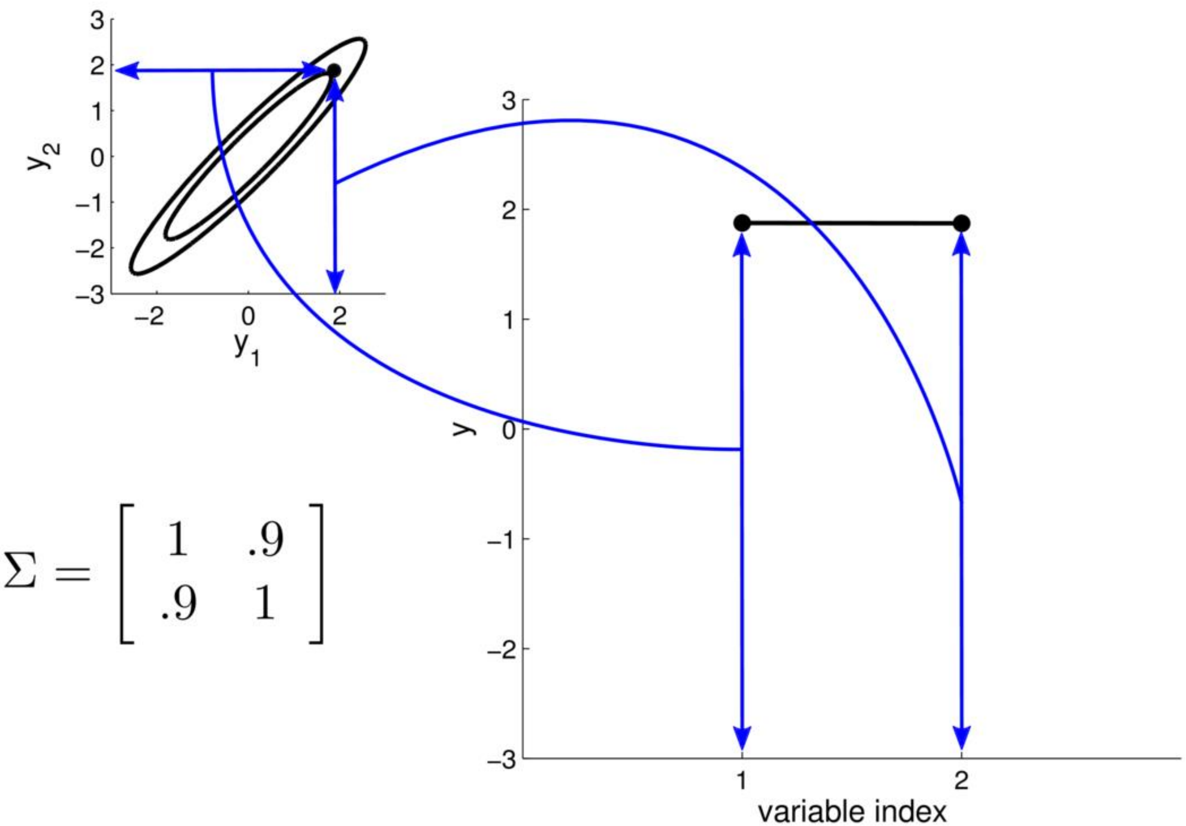
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

Visualizing Gaussian Processes



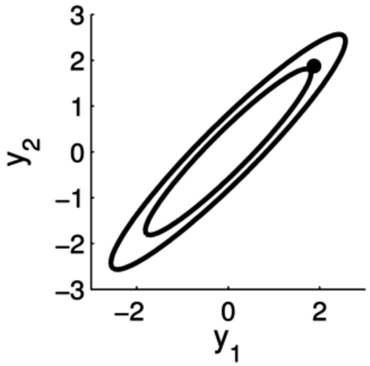
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

Visualizing Gaussian Processes

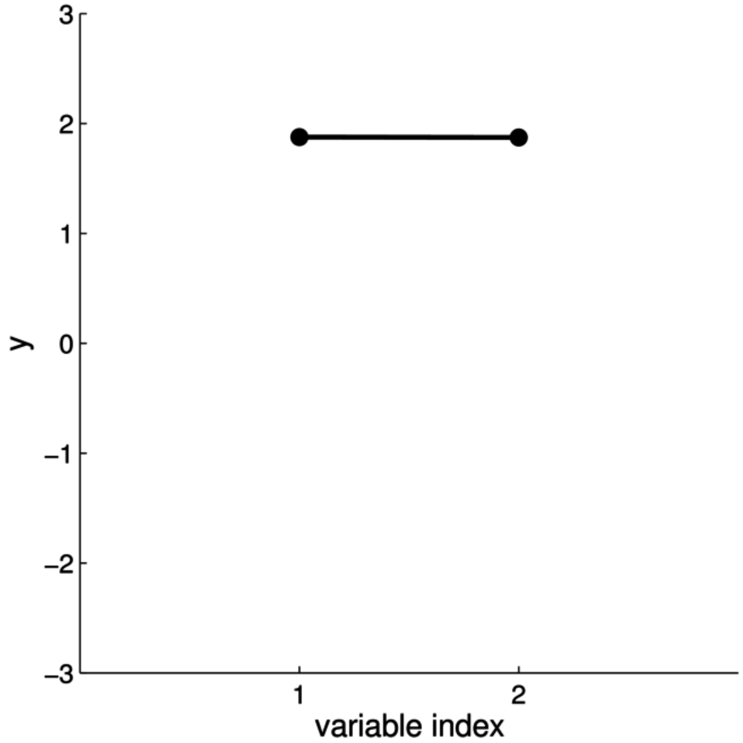


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

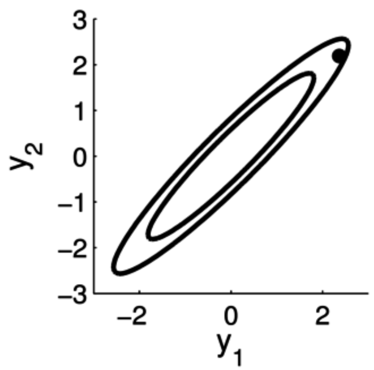
Visualizing Gaussian Processes



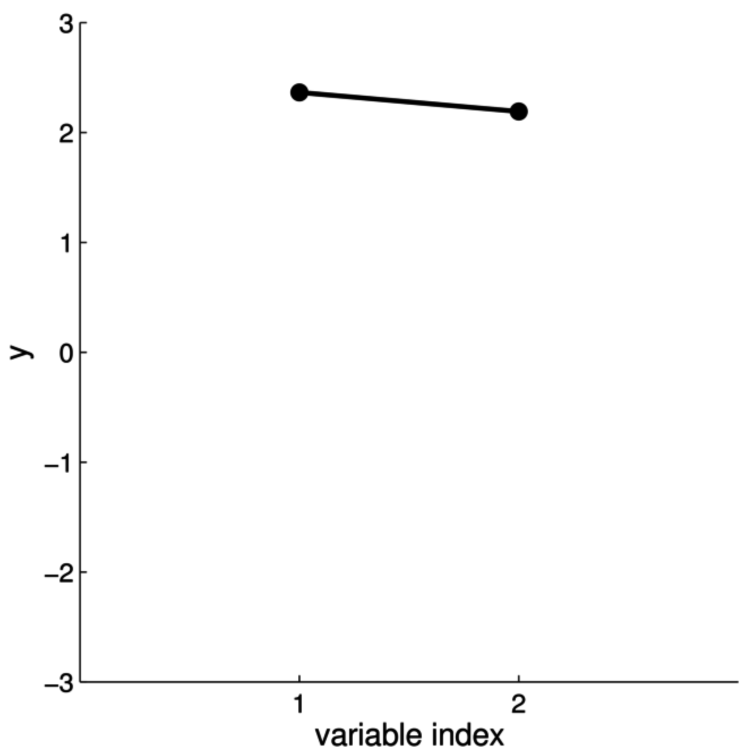
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



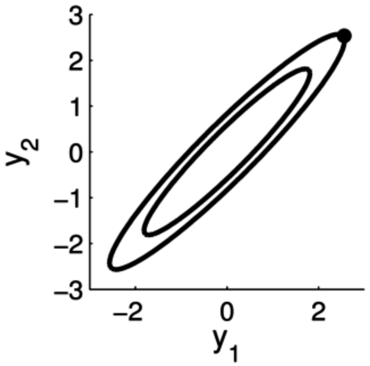
Visualizing Gaussian Processes



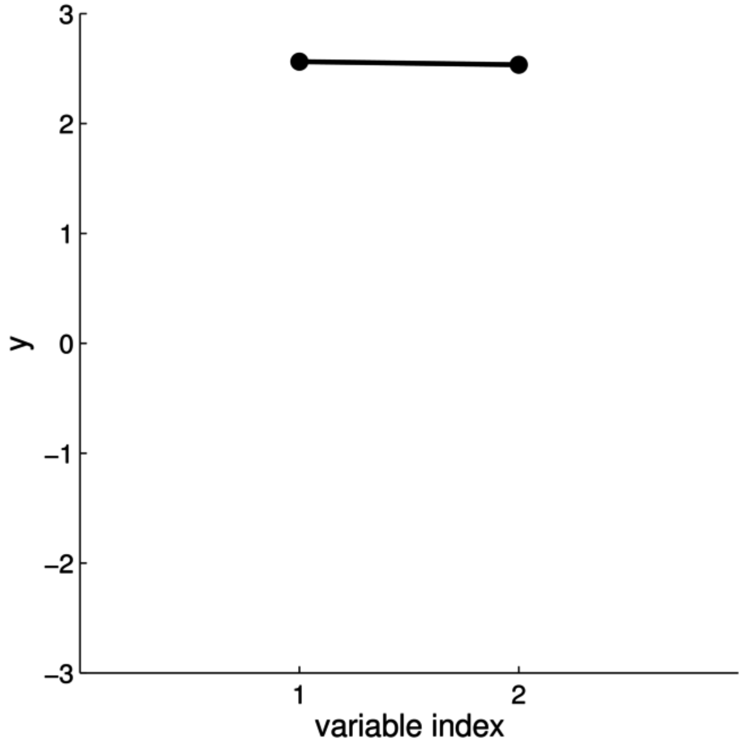
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



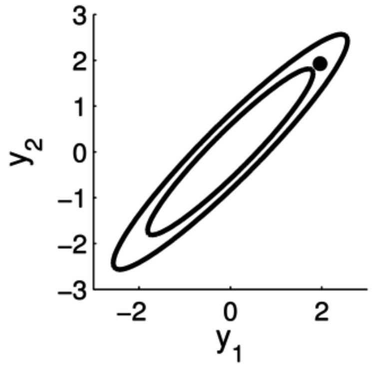
Visualizing Gaussian Processes



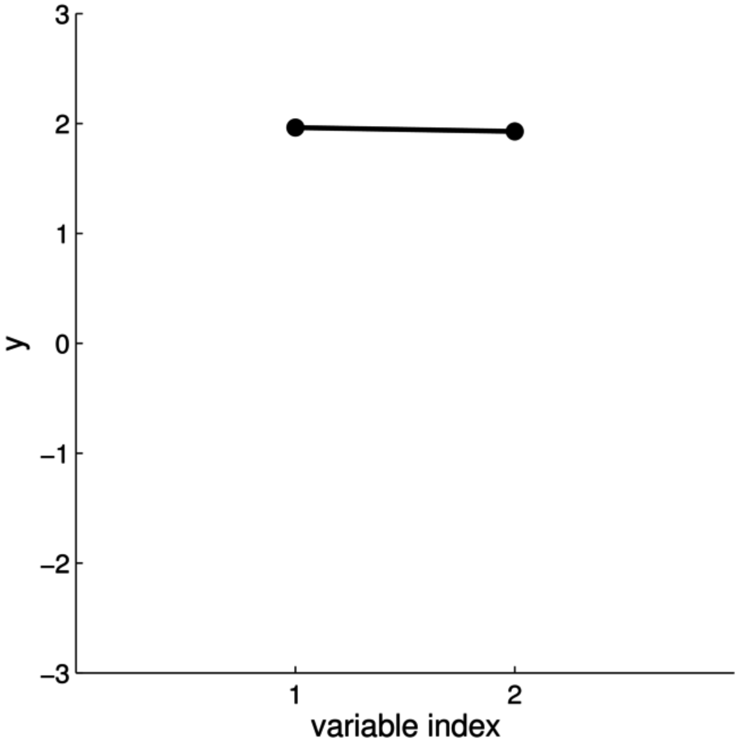
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



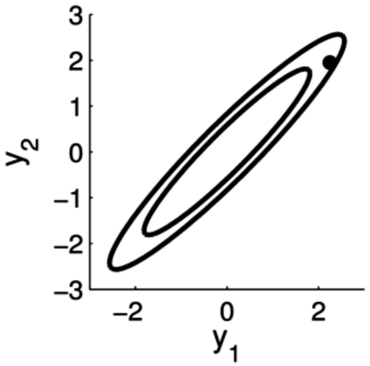
Visualizing Gaussian Processes



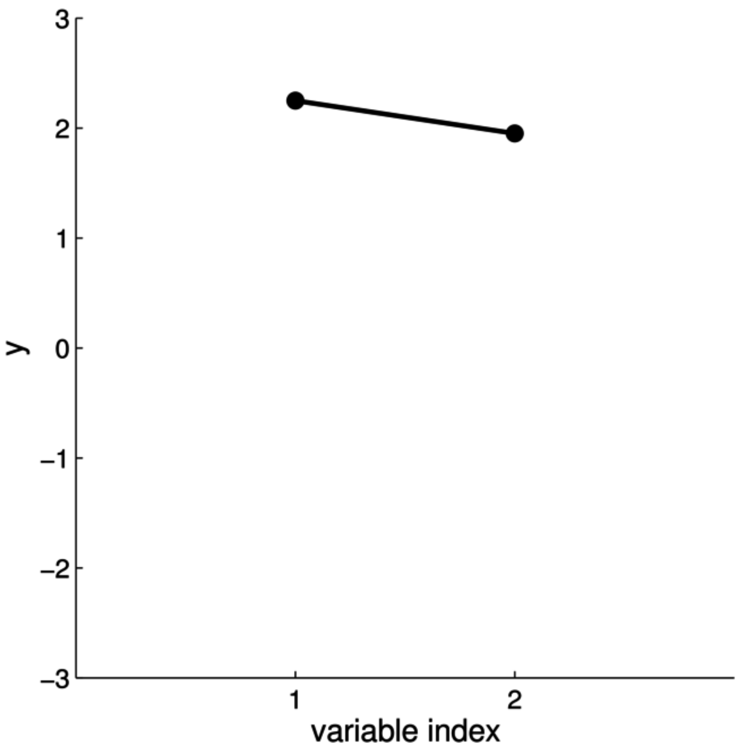
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



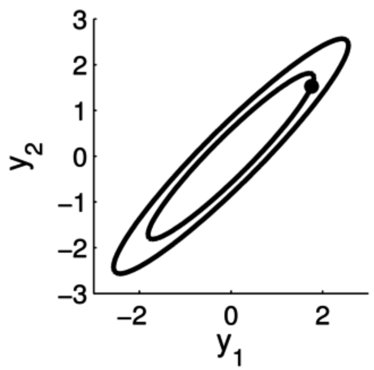
Visualizing Gaussian Processes



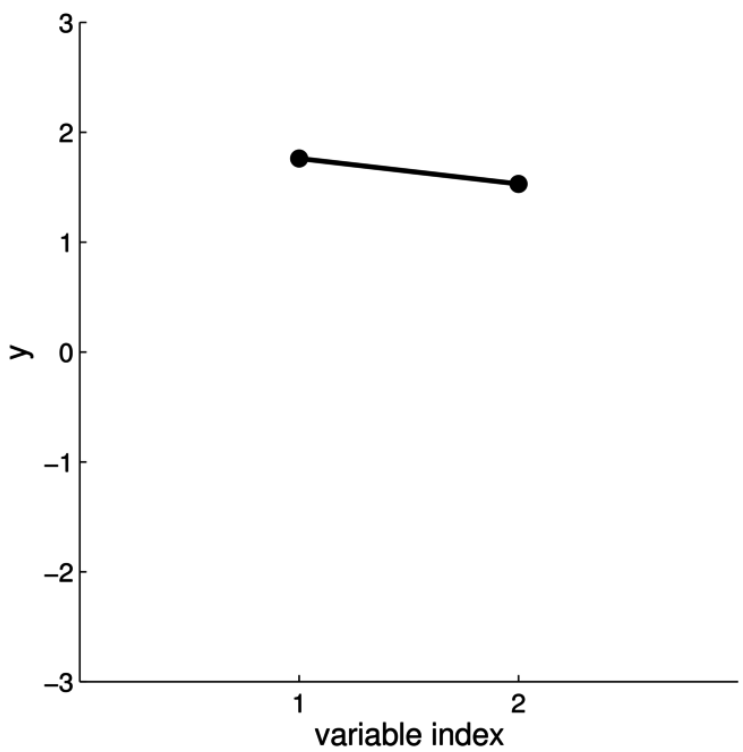
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



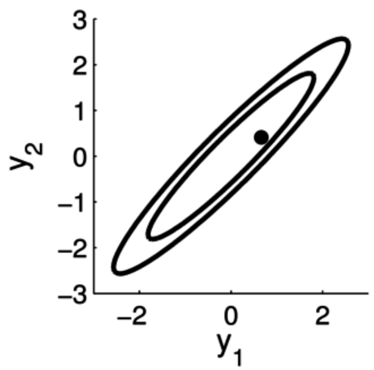
Visualizing Gaussian Processes



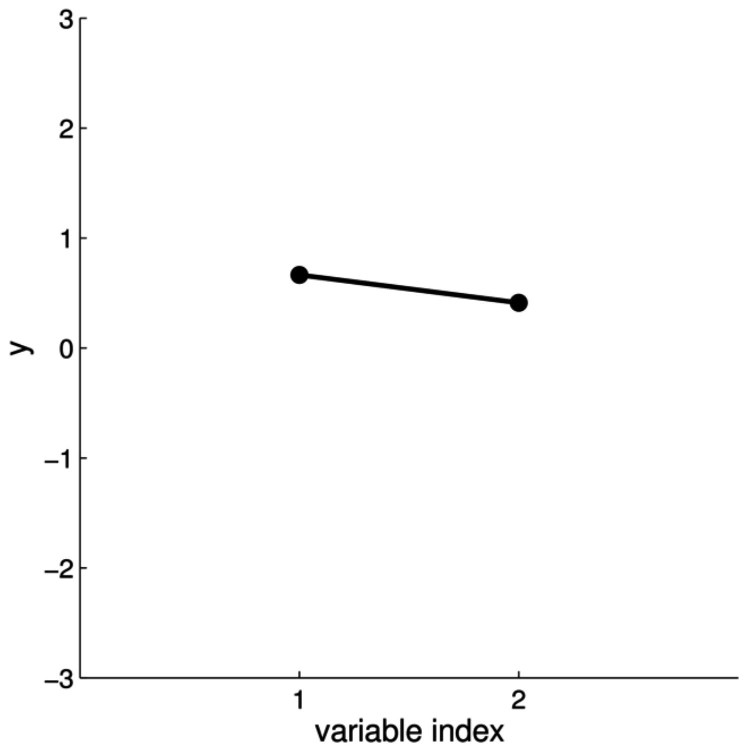
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



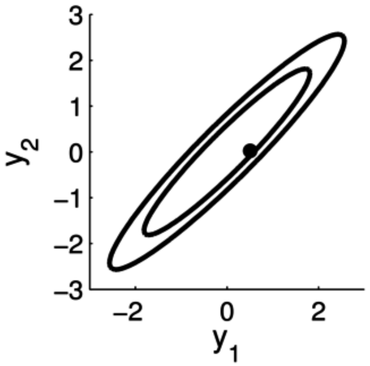
Visualizing Gaussian Processes



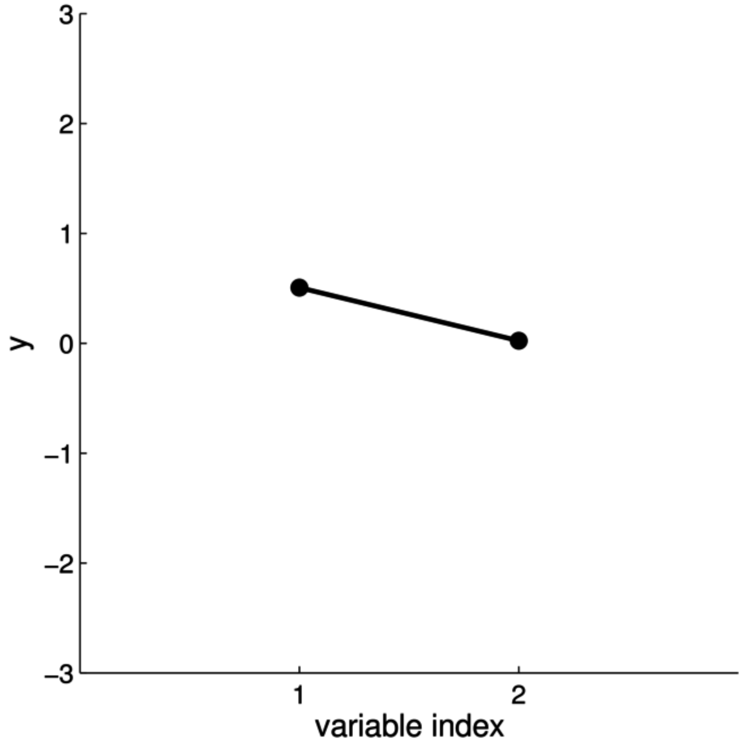
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



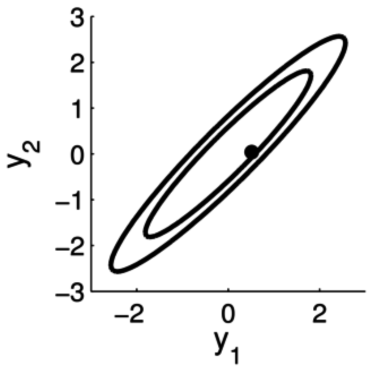
Visualizing Gaussian Processes



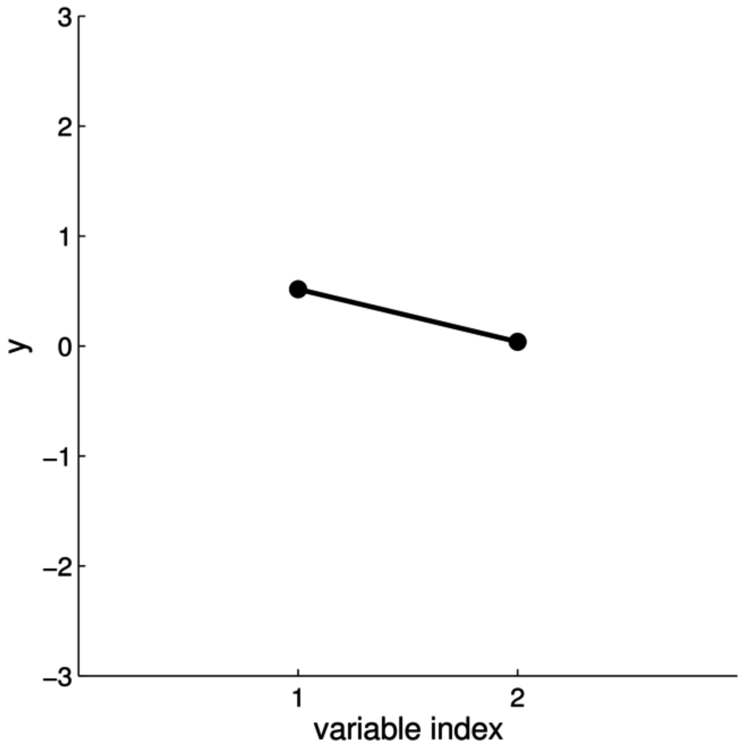
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



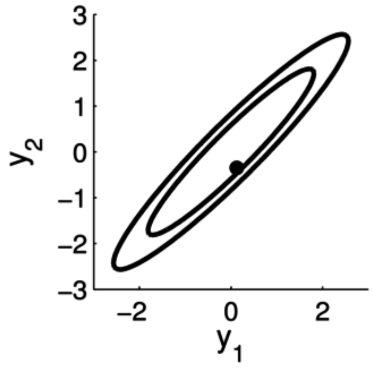
Visualizing Gaussian Processes



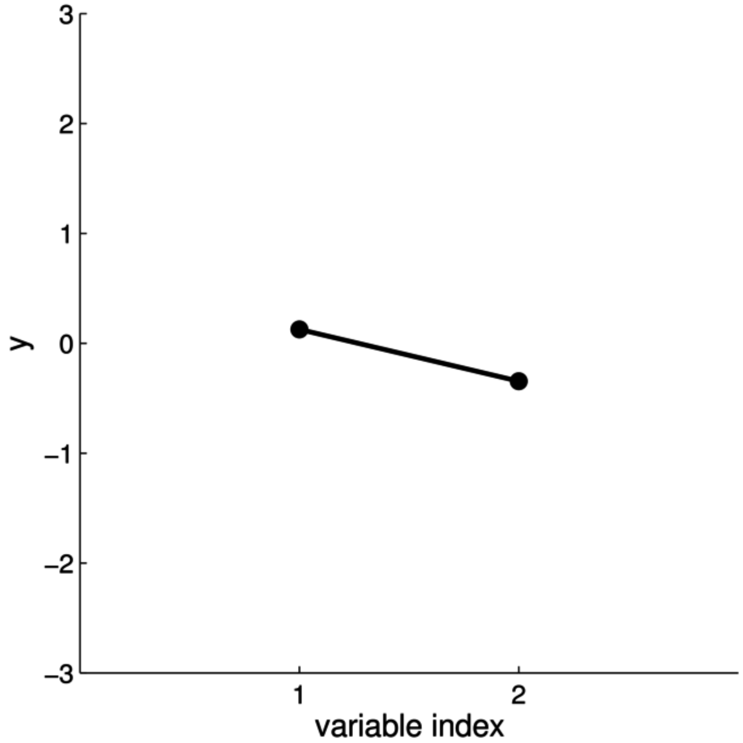
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



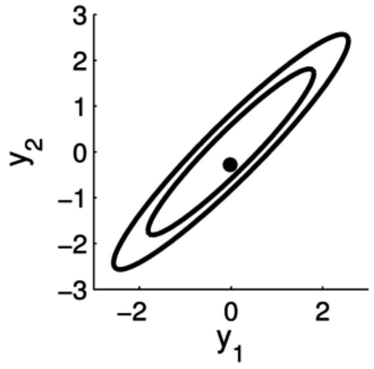
Visualizing Gaussian Processes



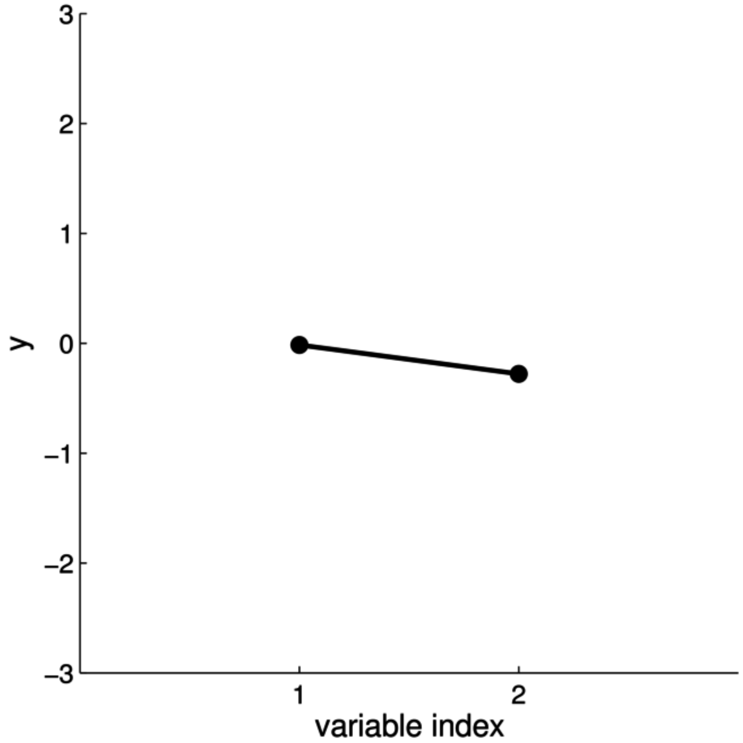
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



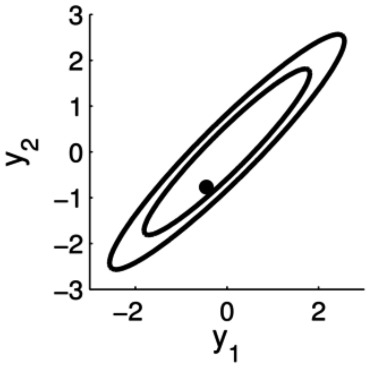
Visualizing Gaussian Processes



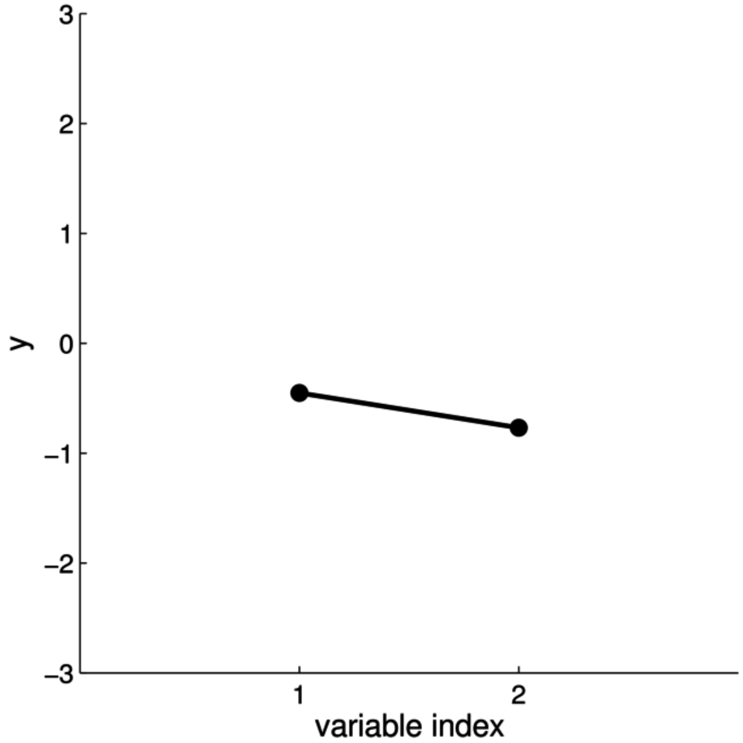
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



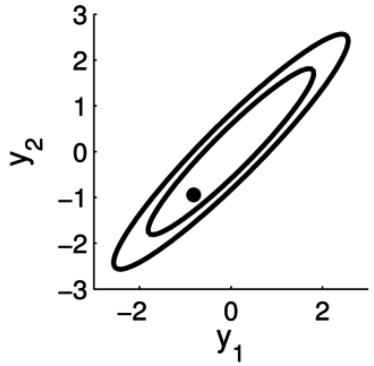
Visualizing Gaussian Processes



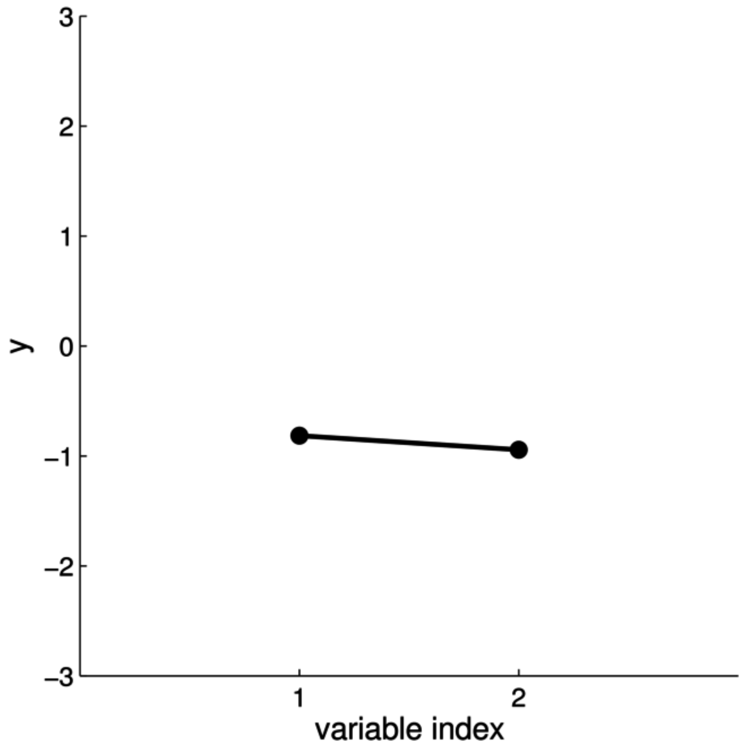
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



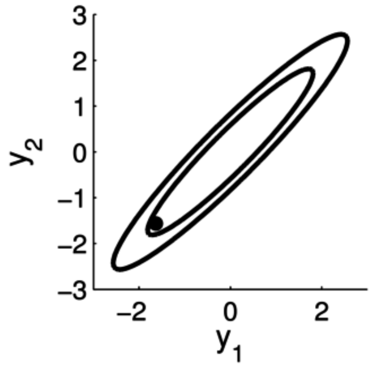
Visualizing Gaussian Processes



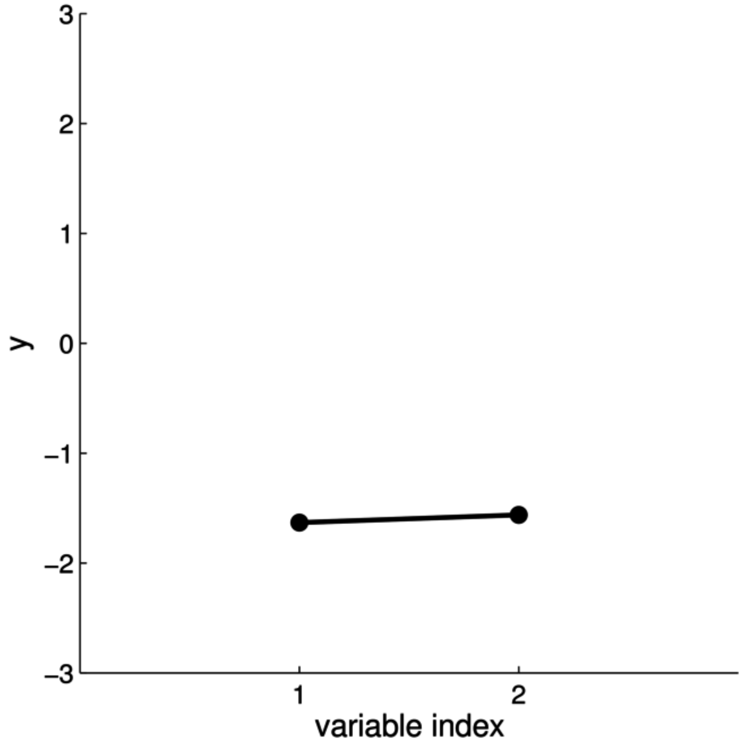
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



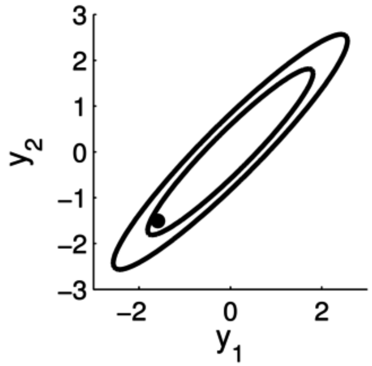
Visualizing Gaussian Processes



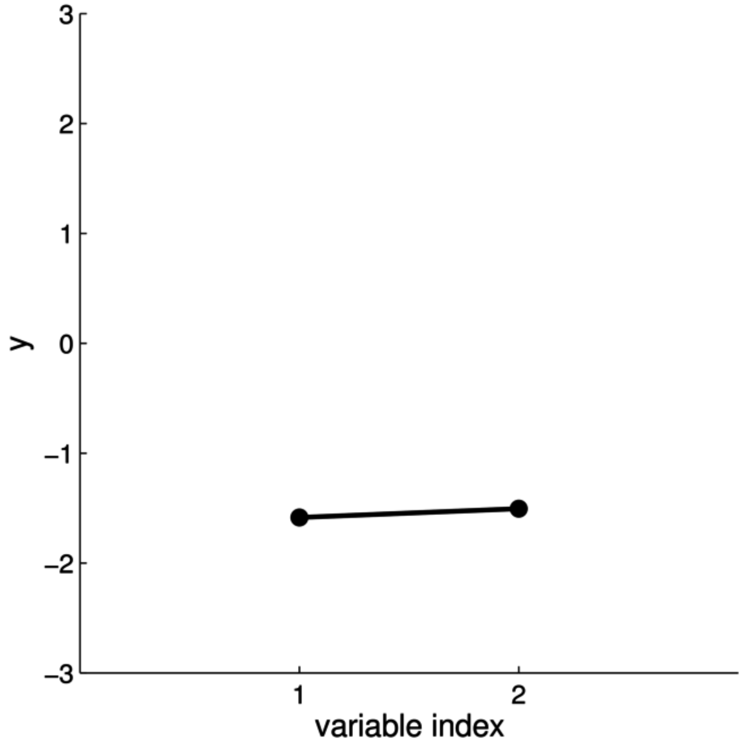
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



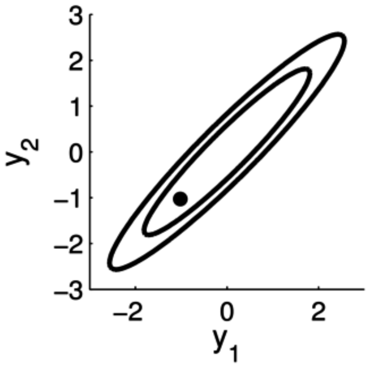
Visualizing Gaussian Processes



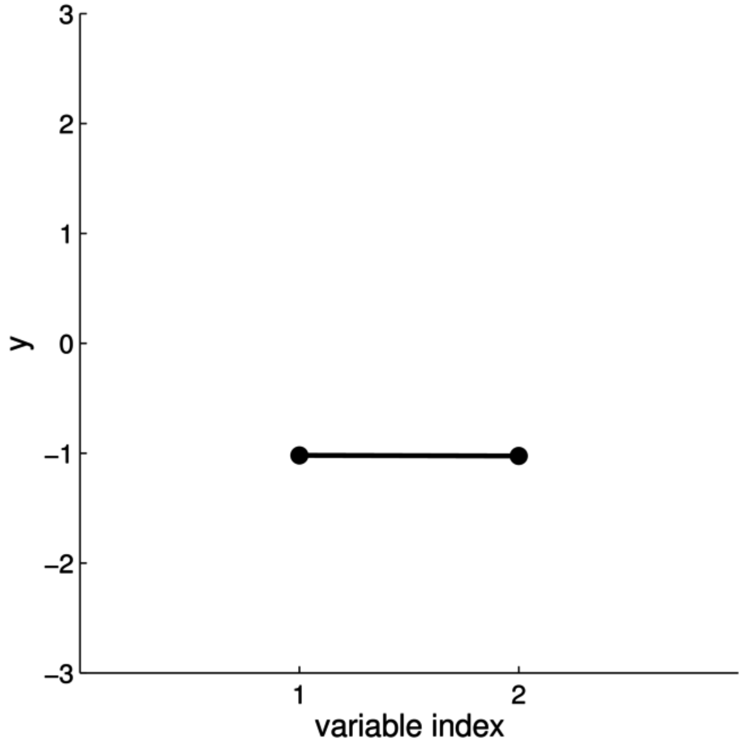
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



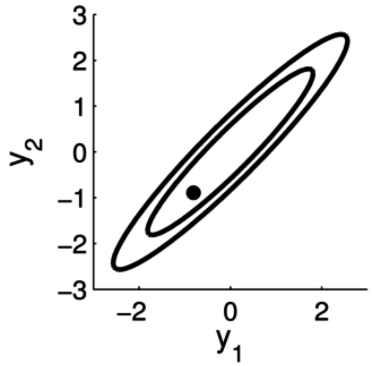
Visualizing Gaussian Processes

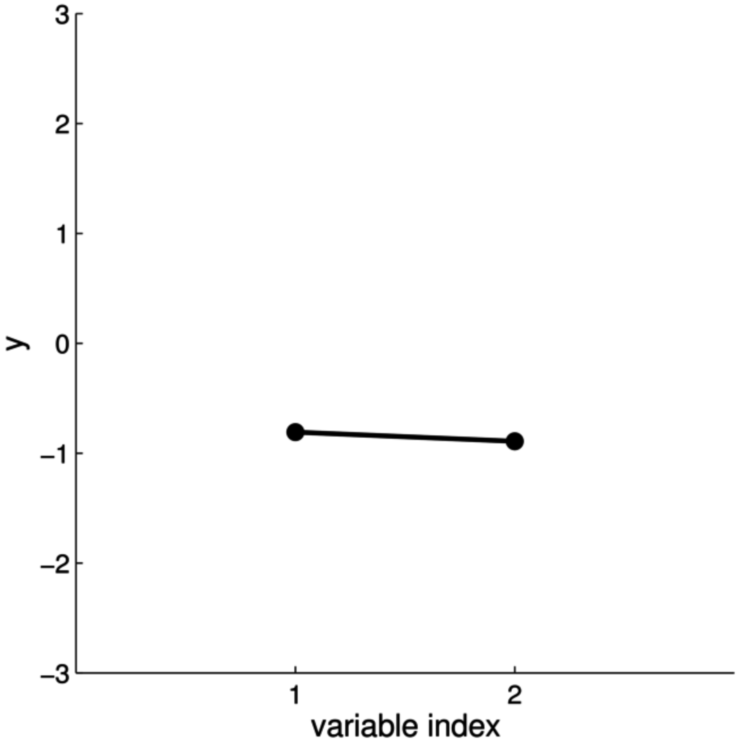


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



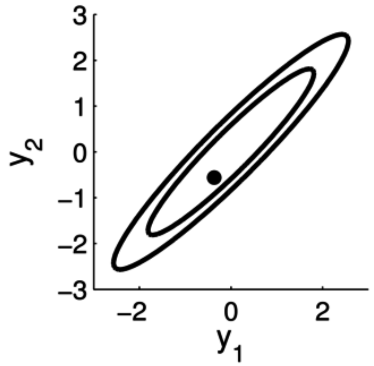
Visualizing Gaussian Processes



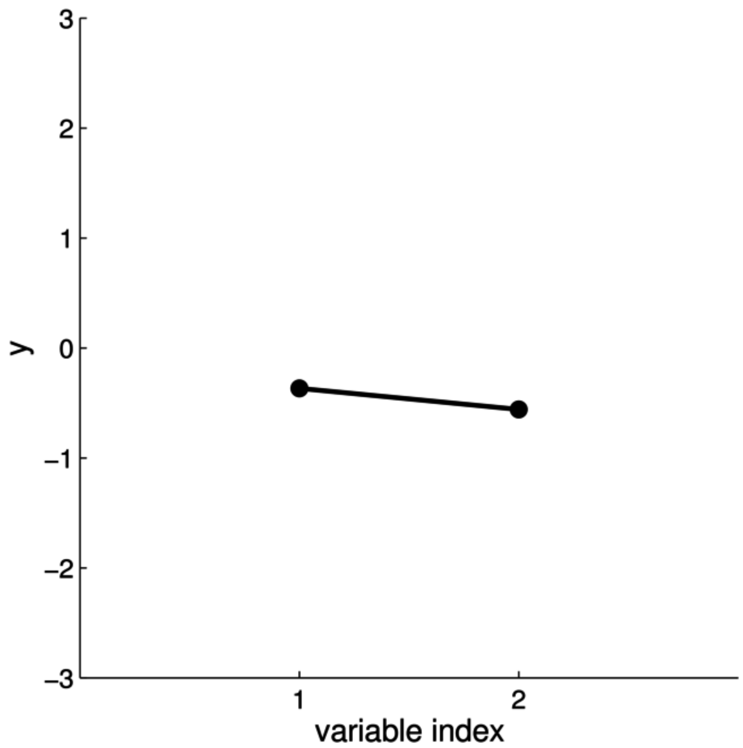


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

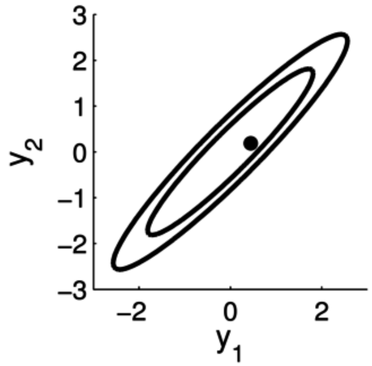
Visualizing Gaussian Processes



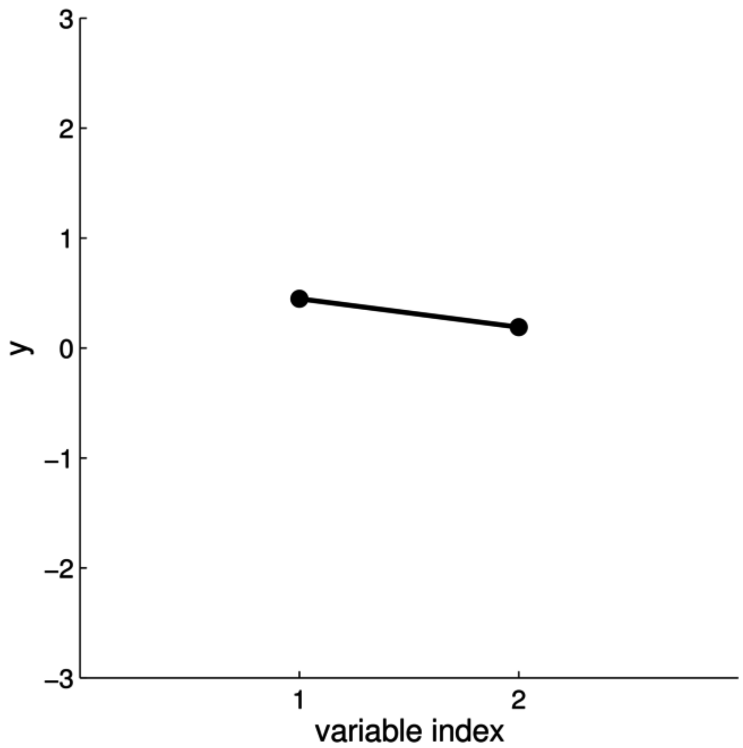
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



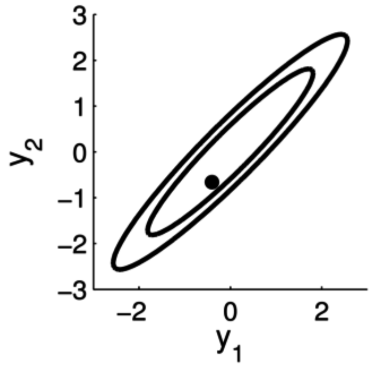
Visualizing Gaussian Processes



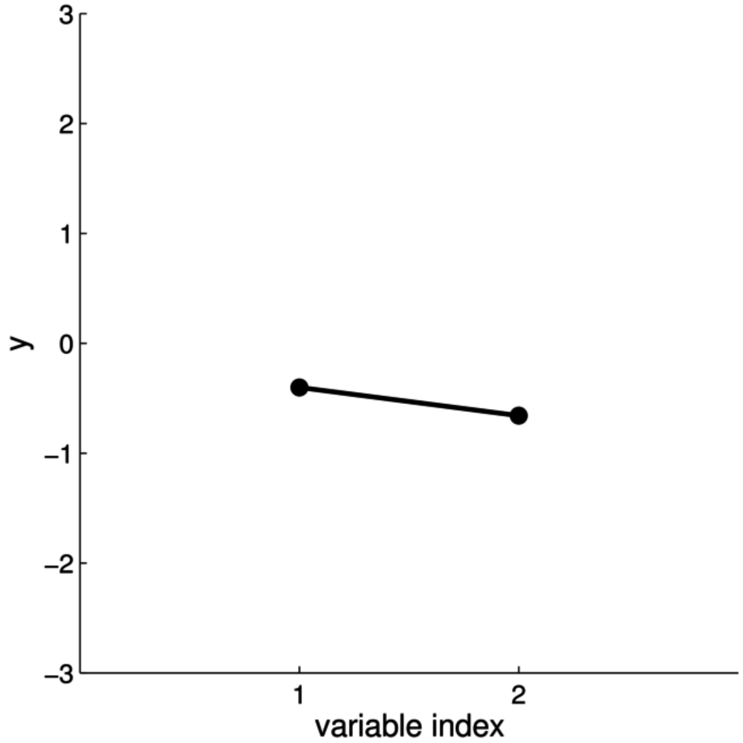
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



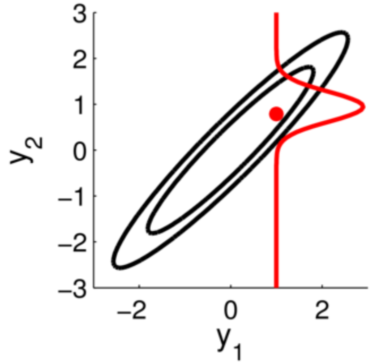
Visualizing Gaussian Processes



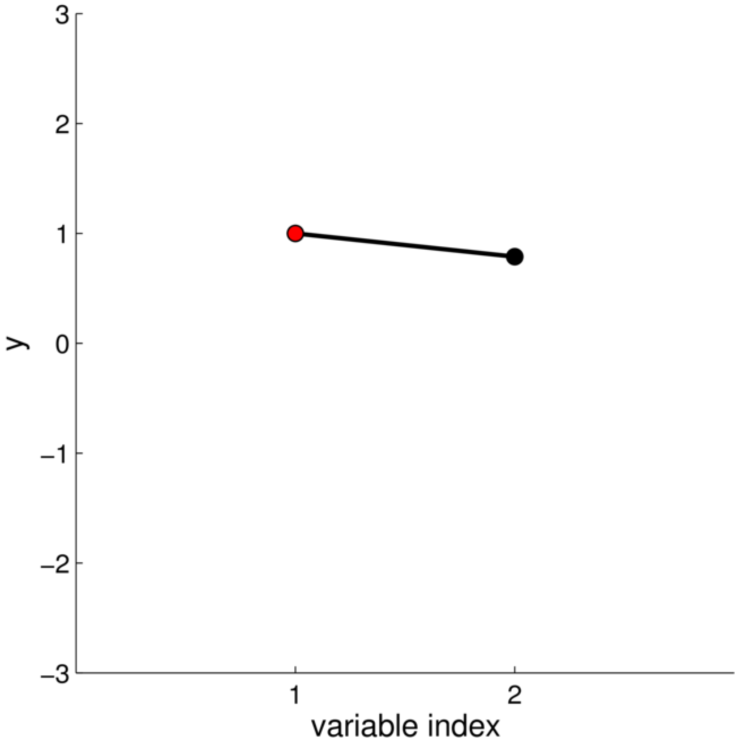
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



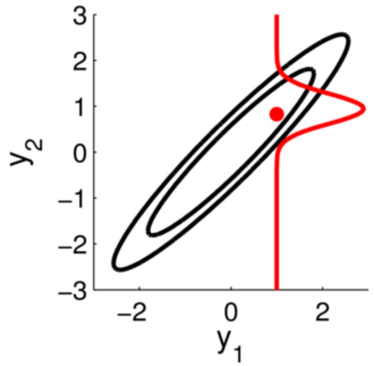
Visualizing Gaussian Processes



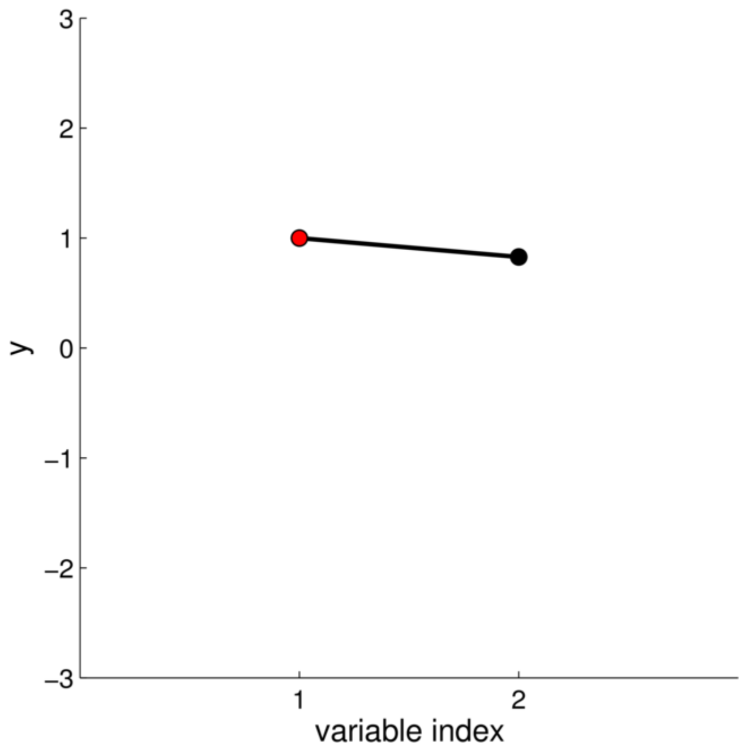
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



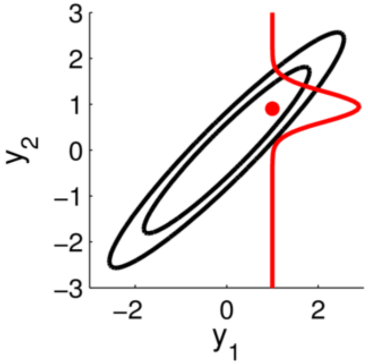
Visualizing Gaussian Processes



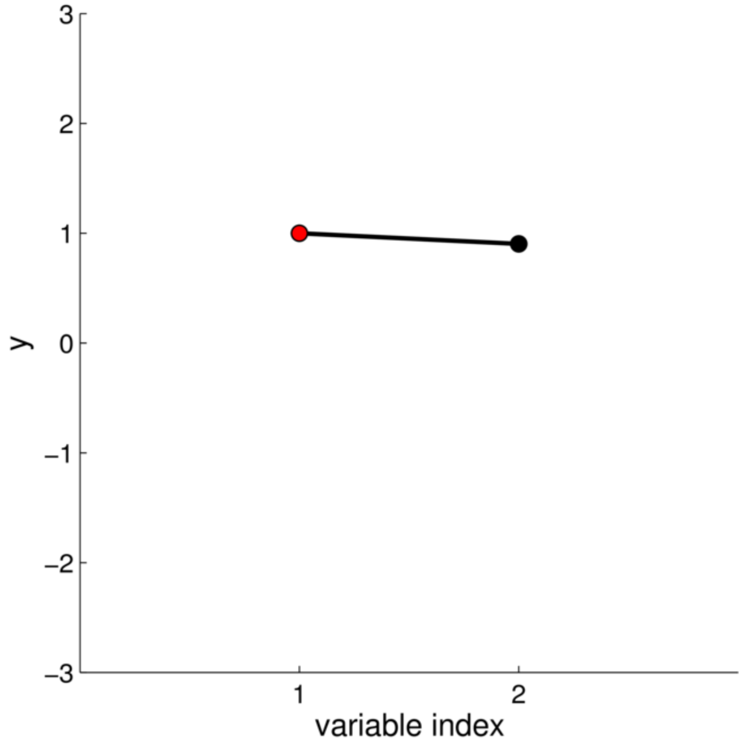
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



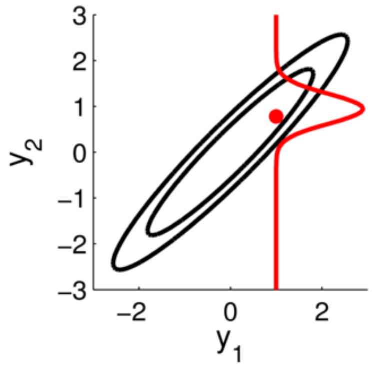
Visualizing Gaussian Processes

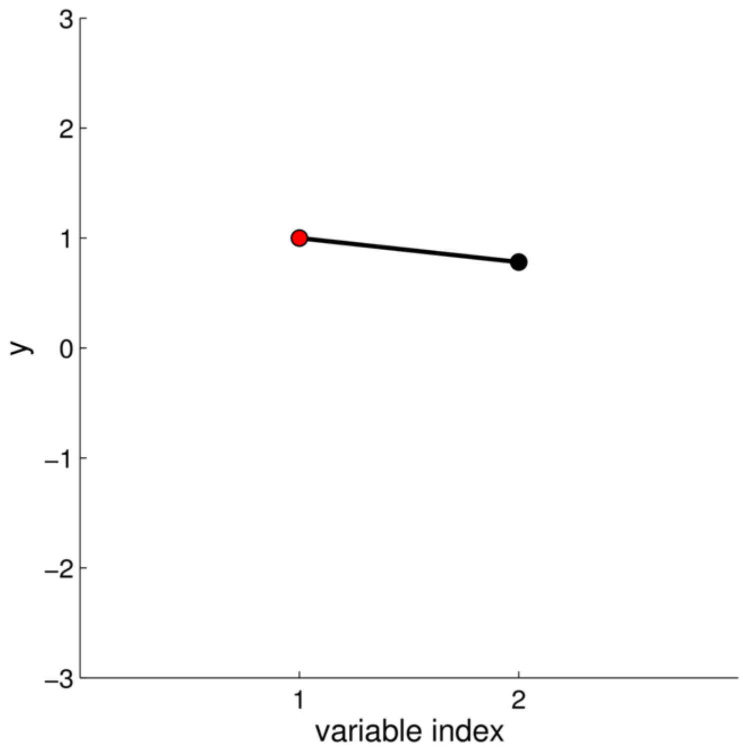


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



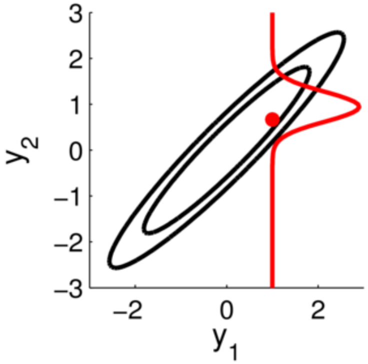
Visualizing Gaussian Processes



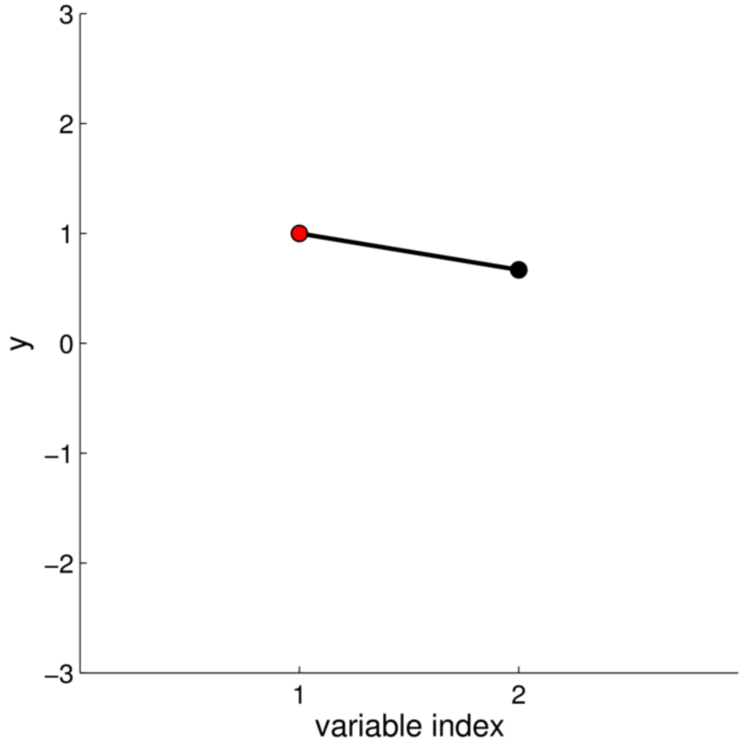


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

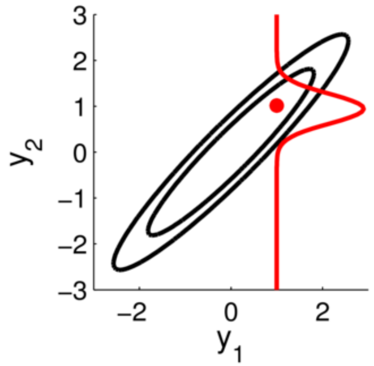
Visualizing Gaussian Processes



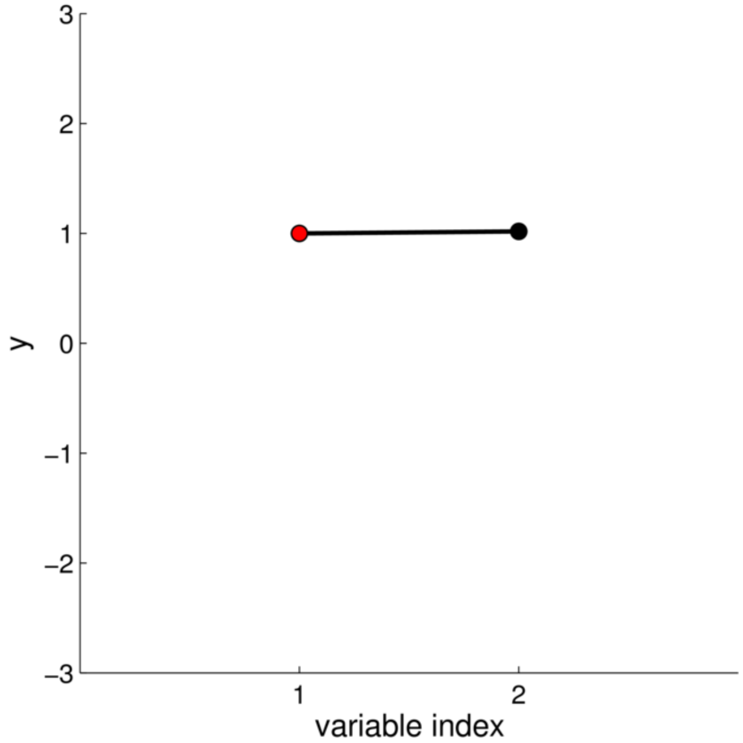
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



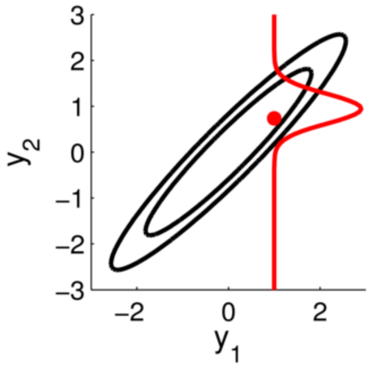
Visualizing Gaussian Processes

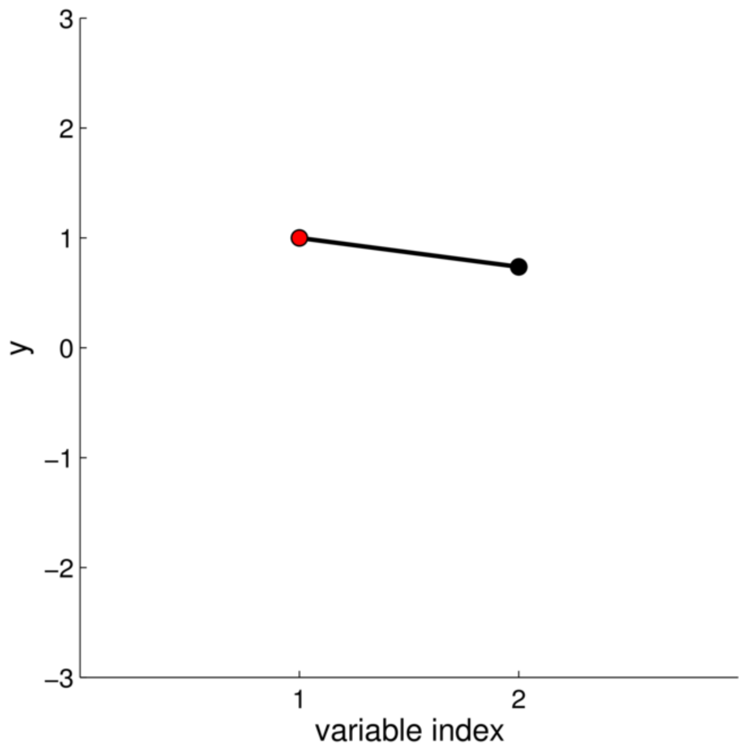


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



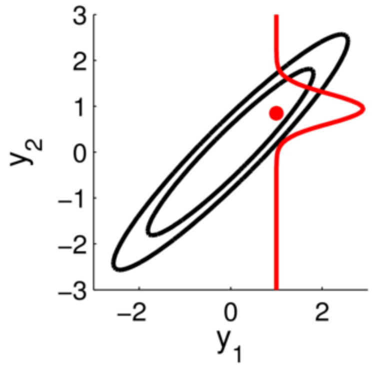
Visualizing Gaussian Processes



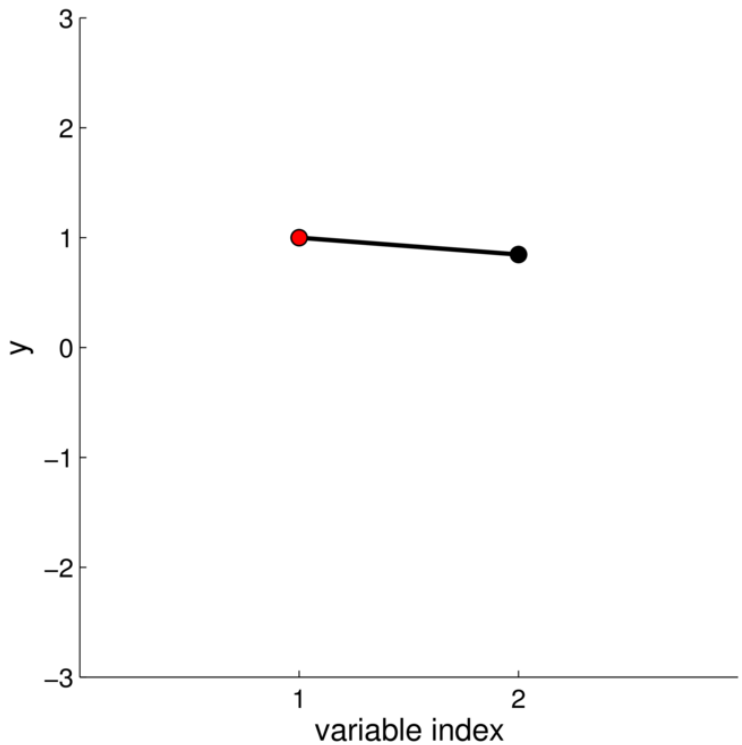


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

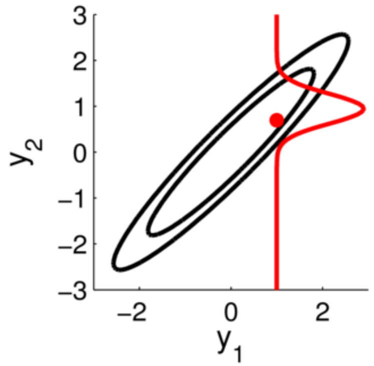
Visualizing Gaussian Processes



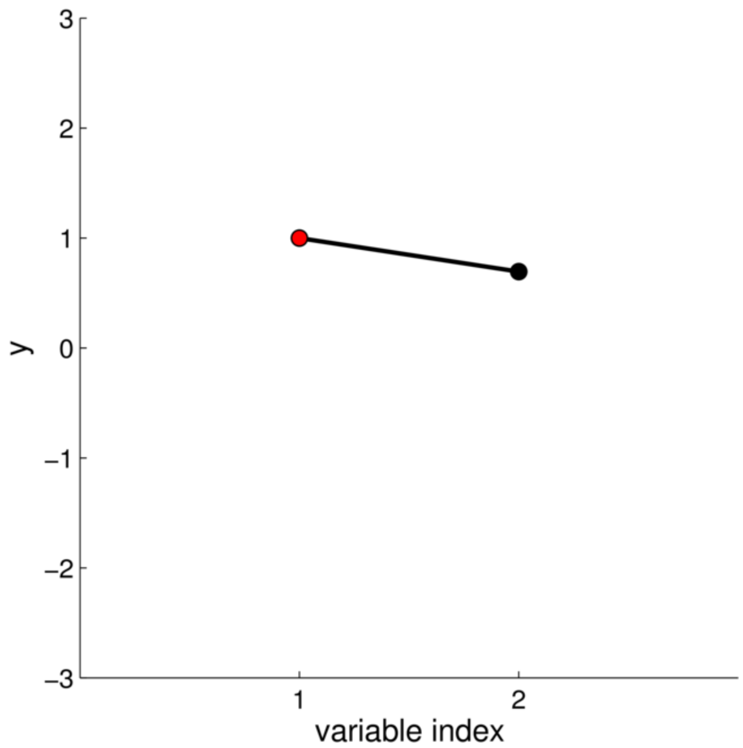
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



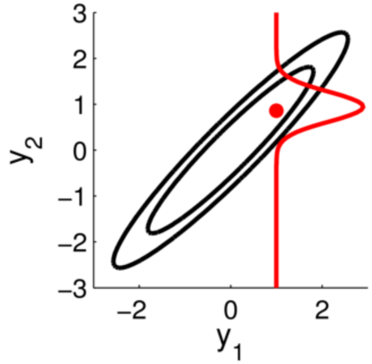
Visualizing Gaussian Processes



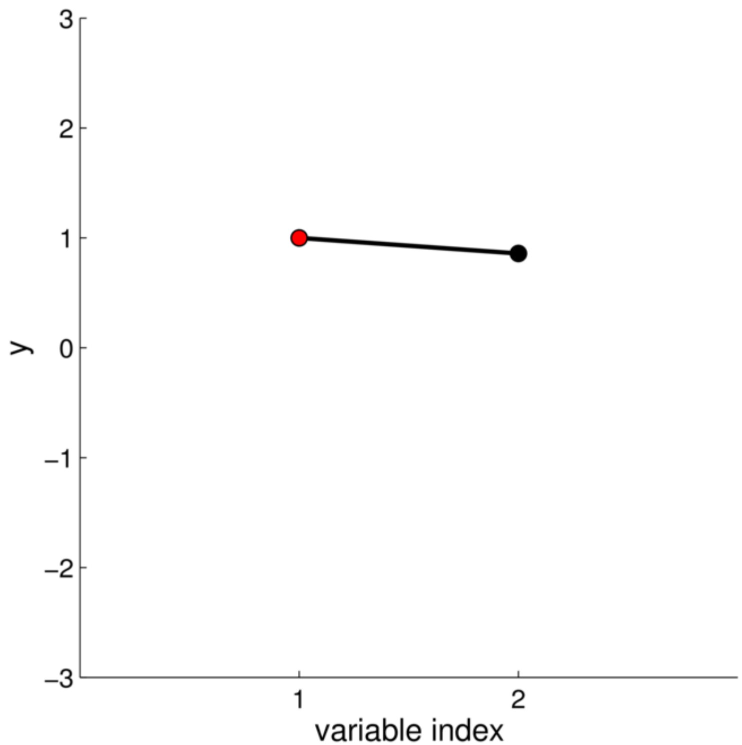
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



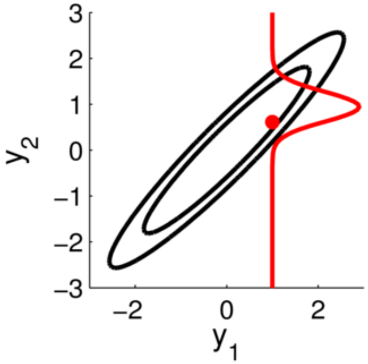
Visualizing Gaussian Processes



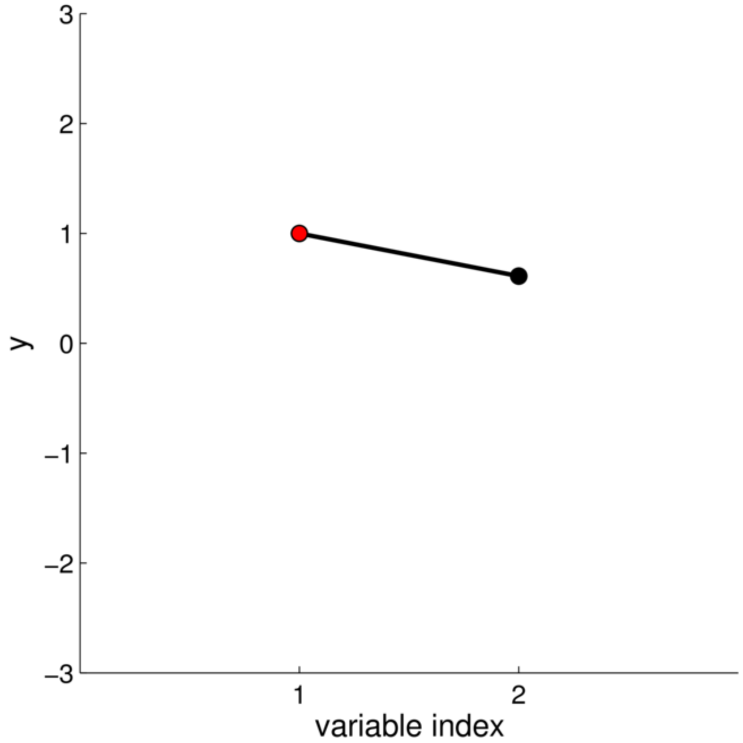
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



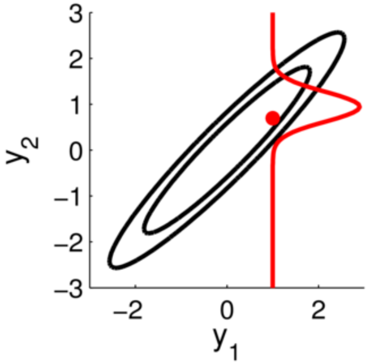
Visualizing Gaussian Processes

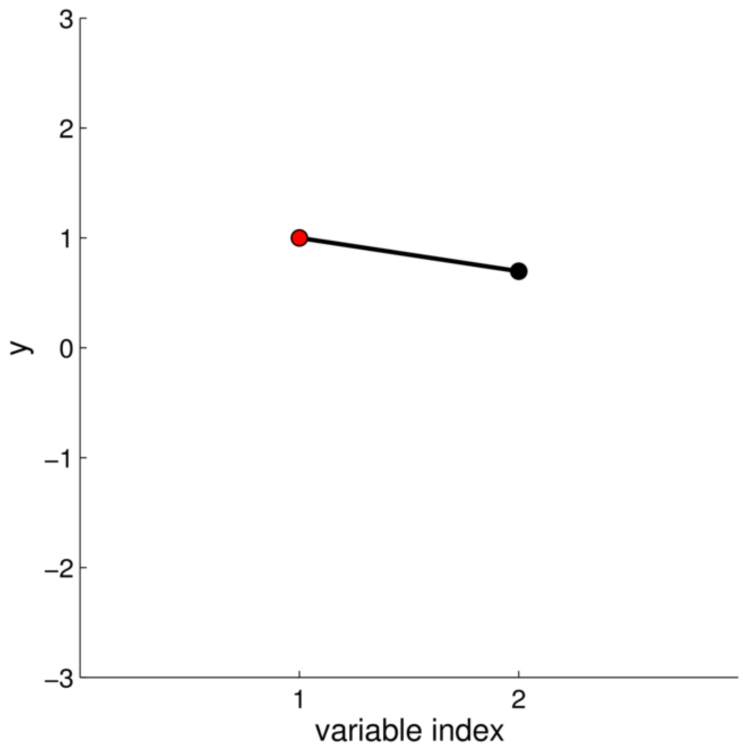


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



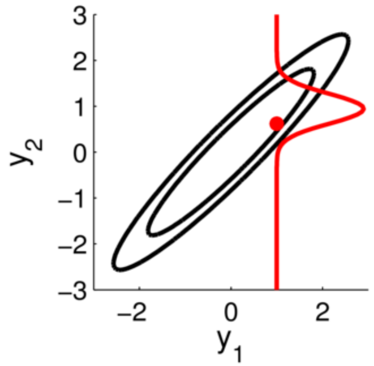
Visualizing Gaussian Processes



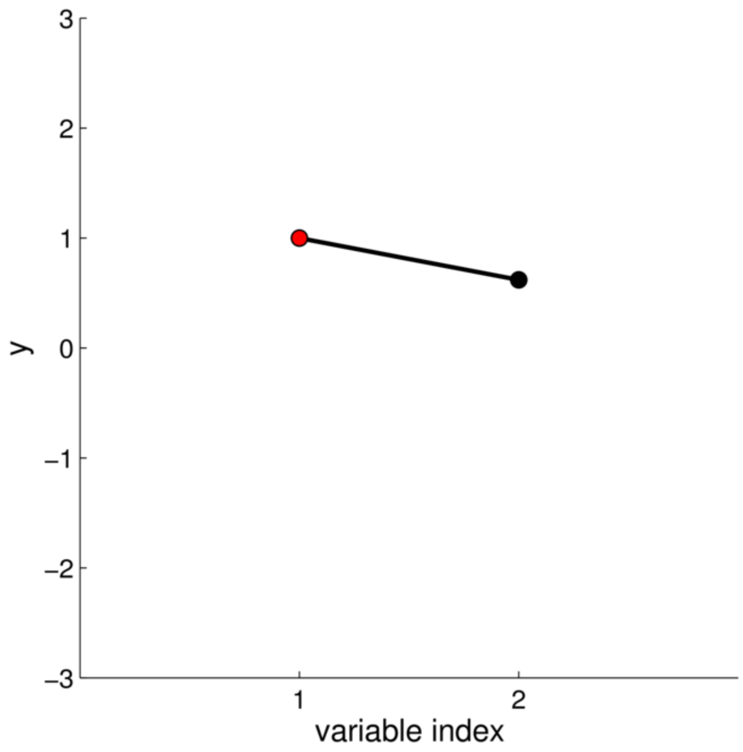


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

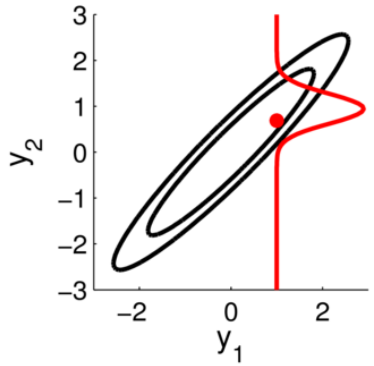
Visualizing Gaussian Processes



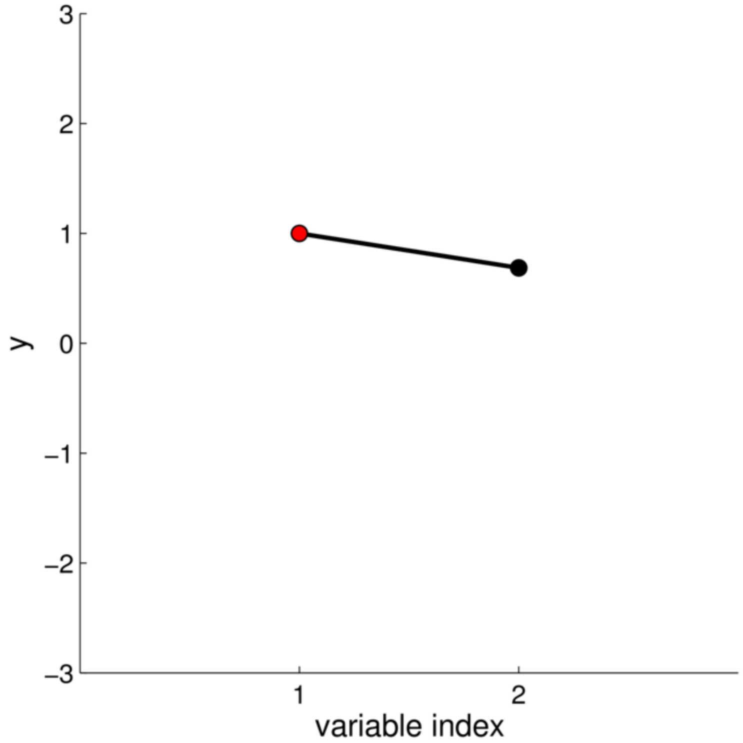
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



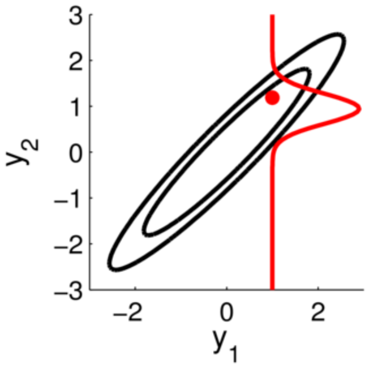
Visualizing Gaussian Processes



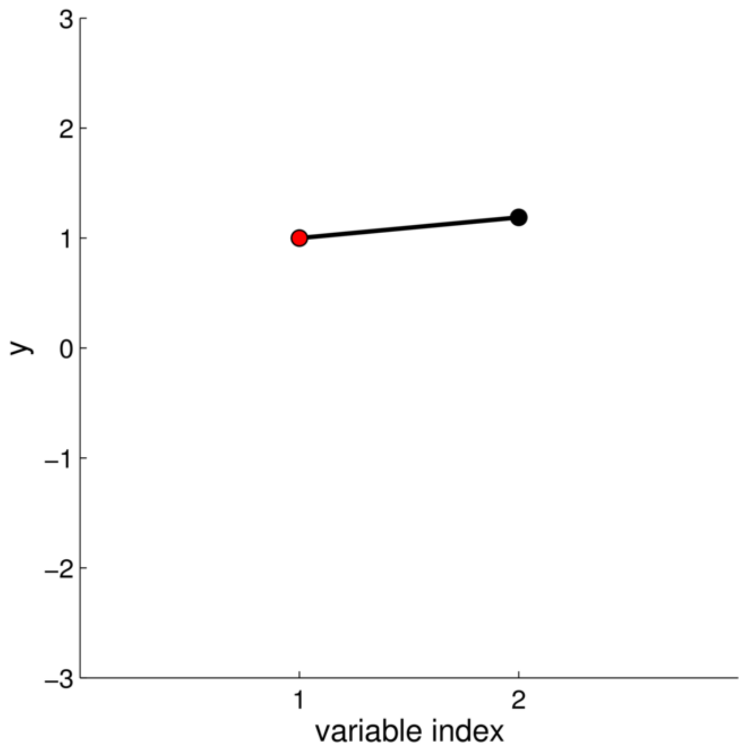
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



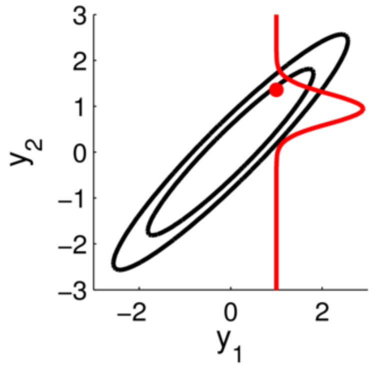
Visualizing Gaussian Processes



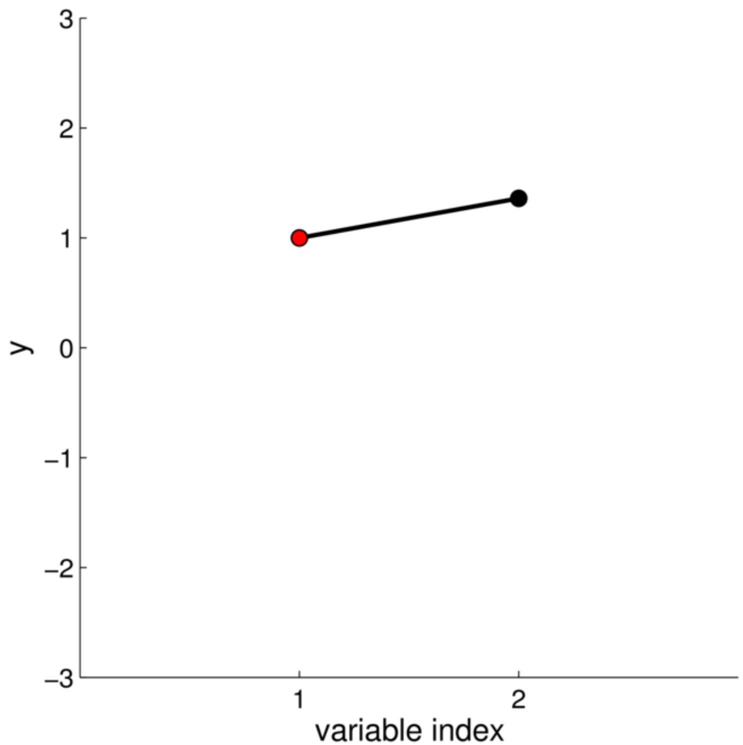
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



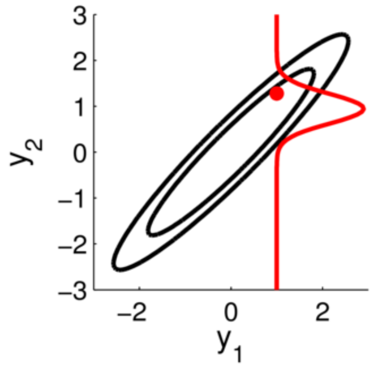
Visualizing Gaussian Processes



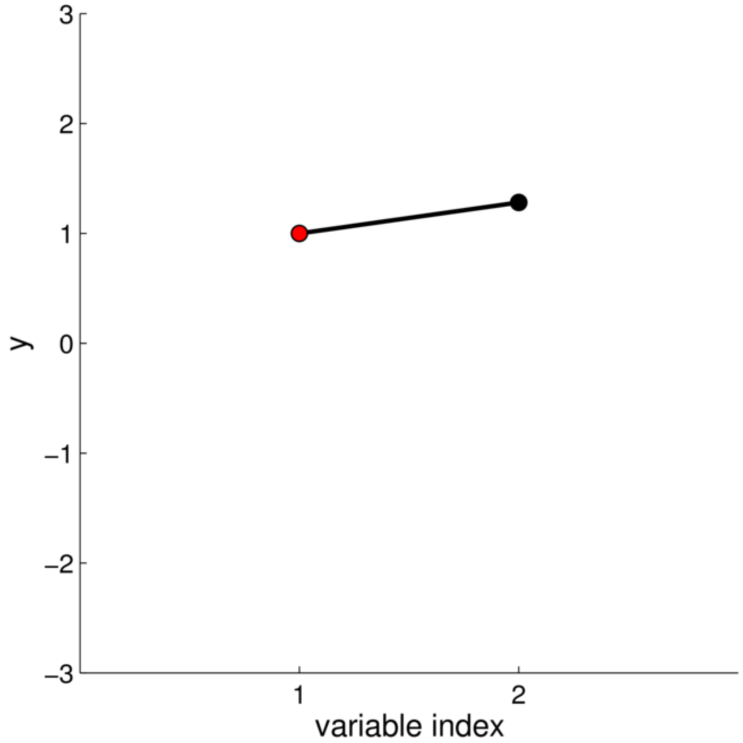
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



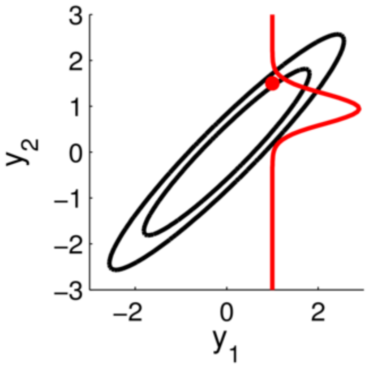
Visualizing Gaussian Processes



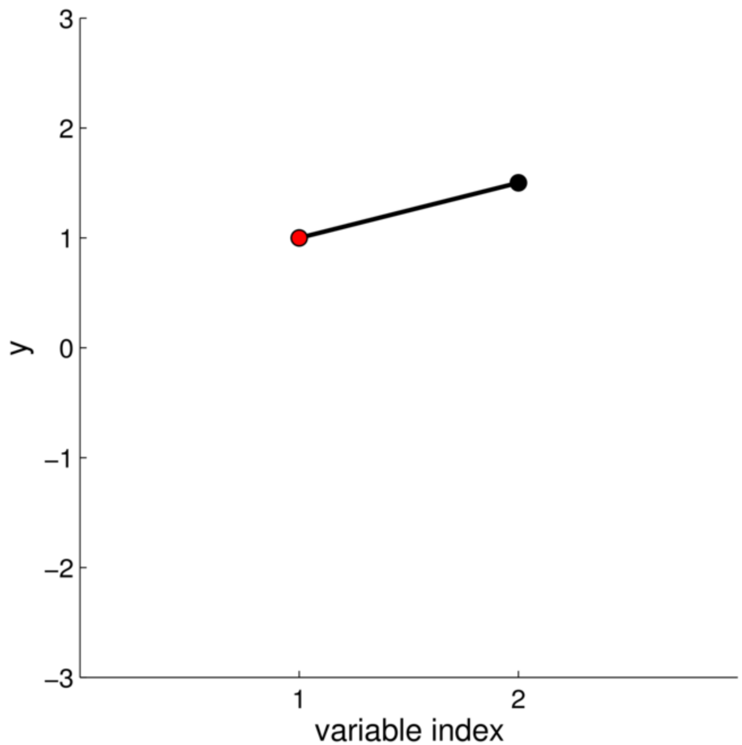
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



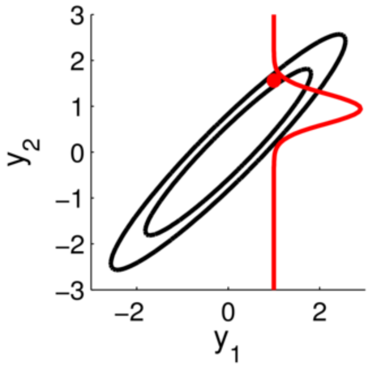
Visualizing Gaussian Processes



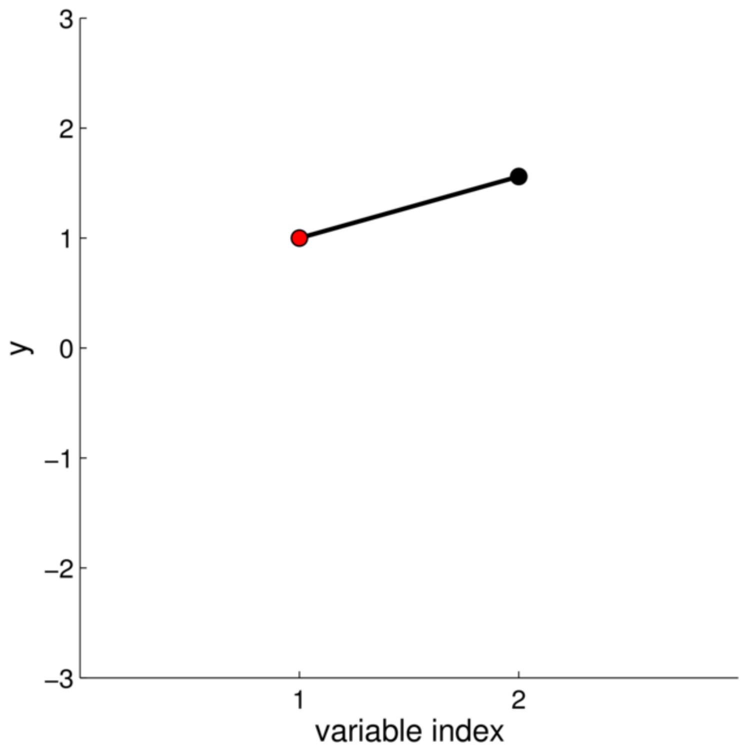
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



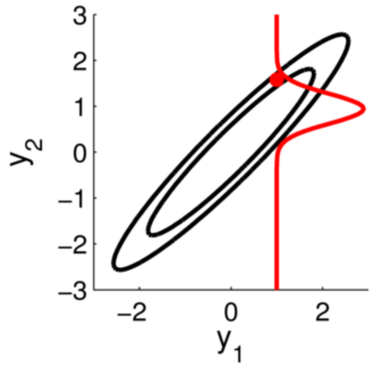
Visualizing Gaussian Processes



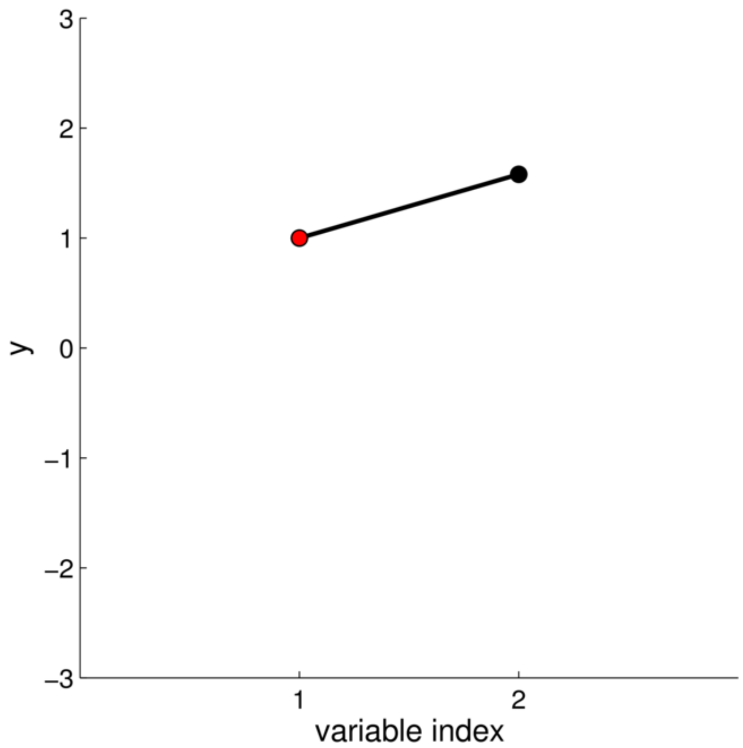
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



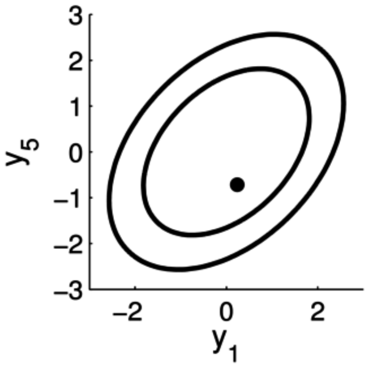
Visualizing Gaussian Processes



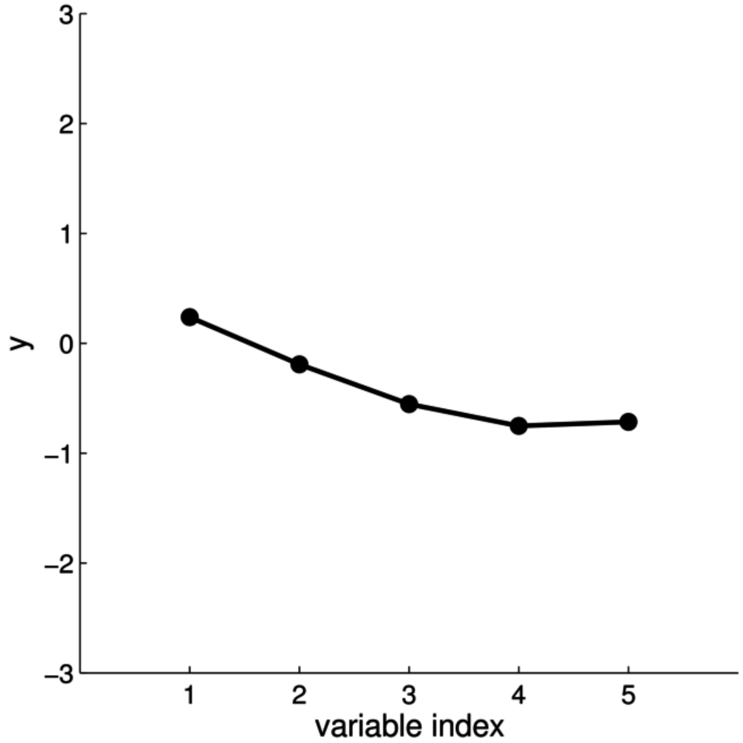
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



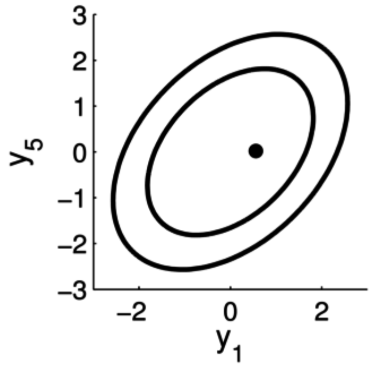
Visualizing Gaussian Processes

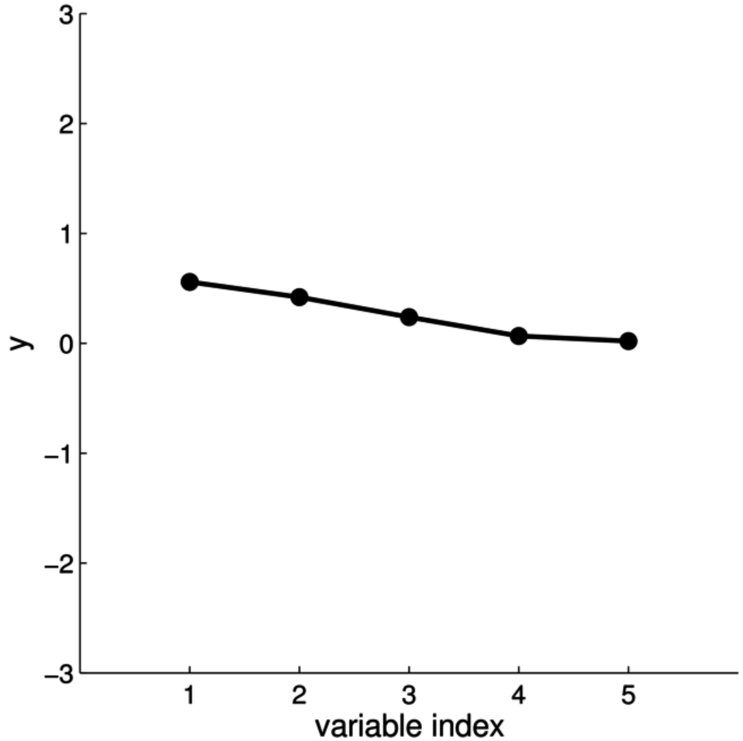


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



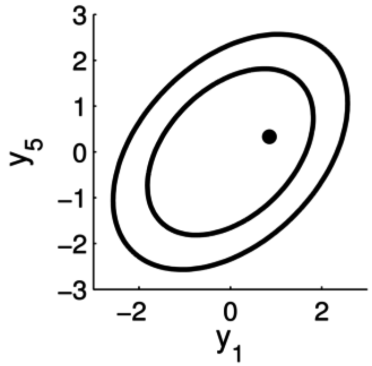
Visualizing Gaussian Processes

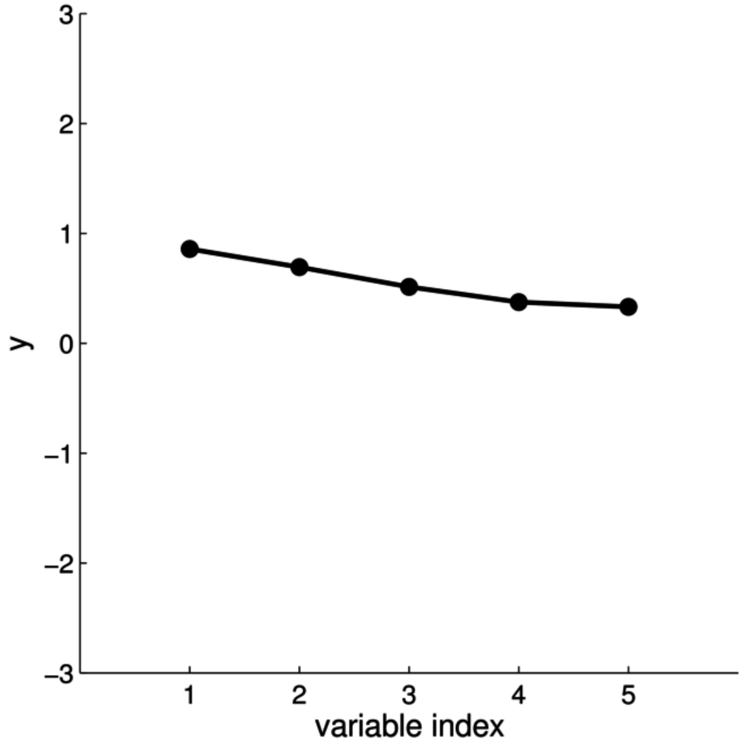




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

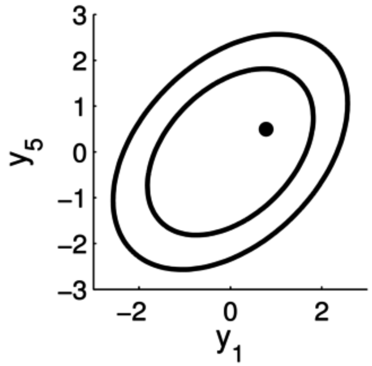
Visualizing Gaussian Processes



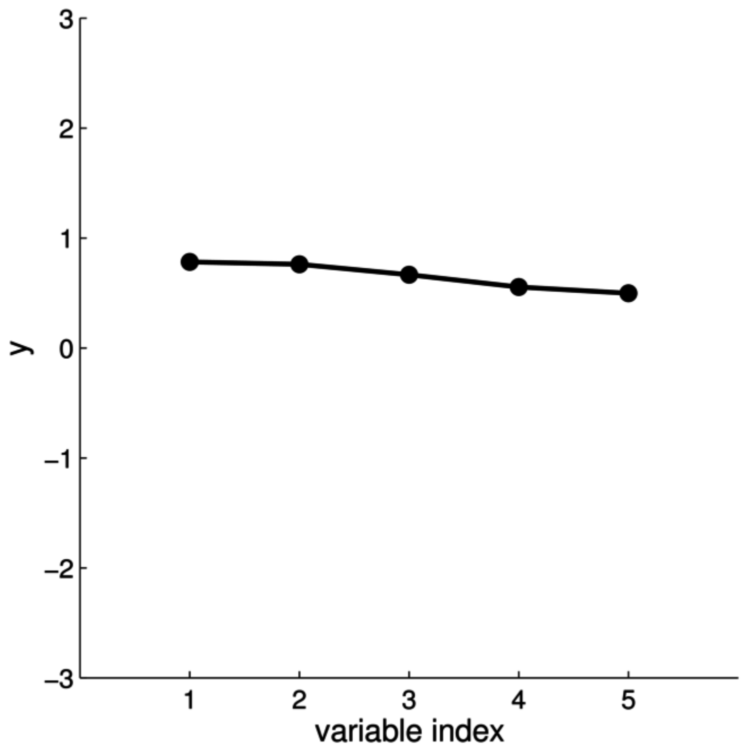


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

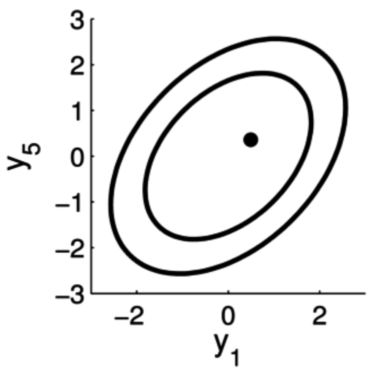
Visualizing Gaussian Processes

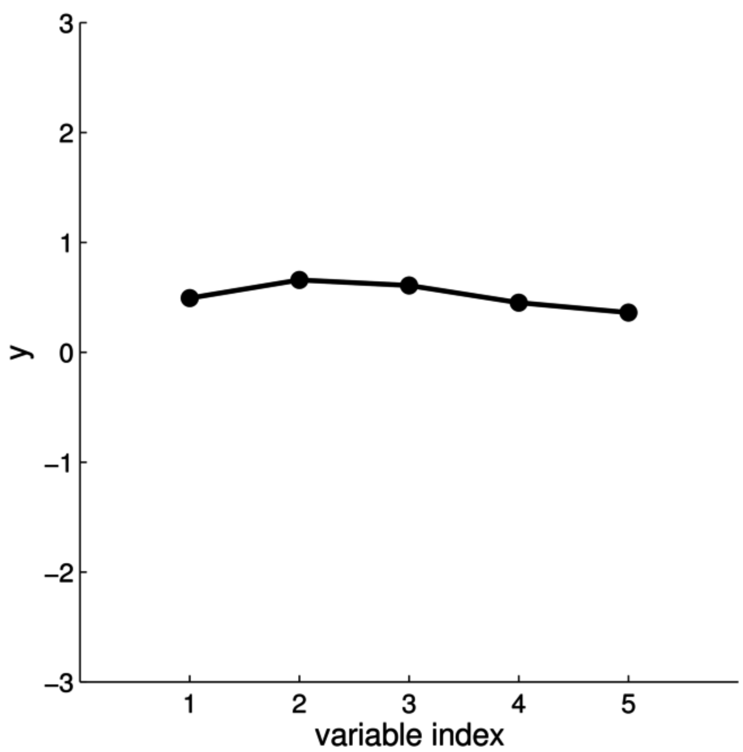


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



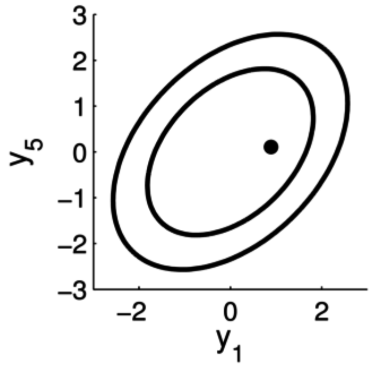
Visualizing Gaussian Processes



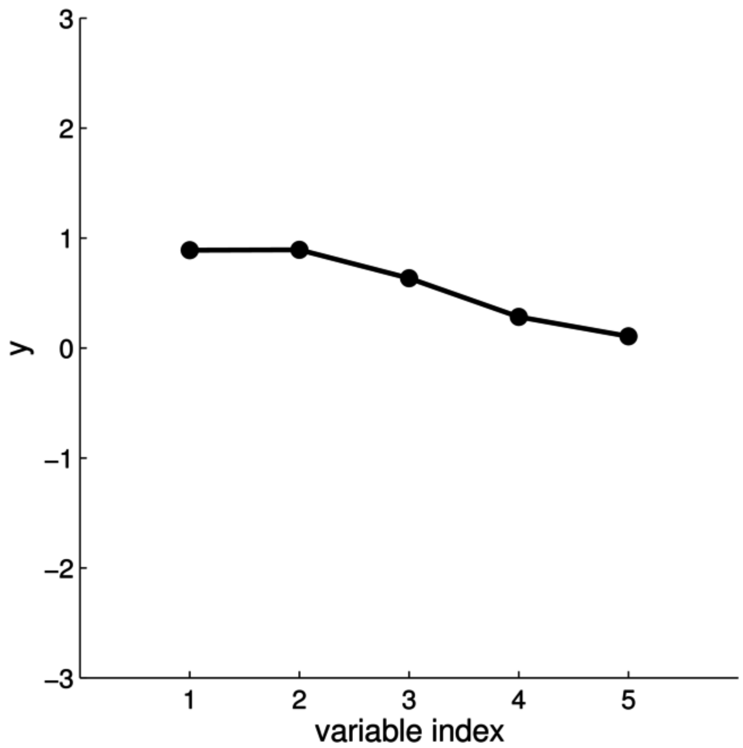


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

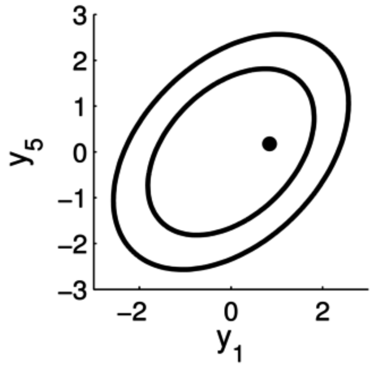
Visualizing Gaussian Processes

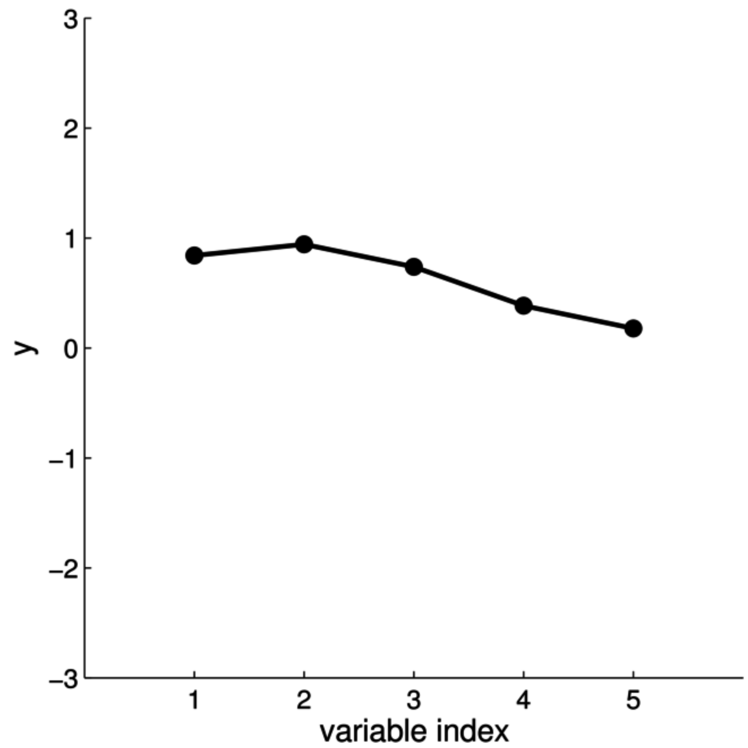


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



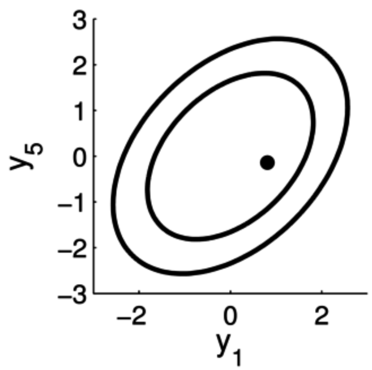
Visualizing Gaussian Processes

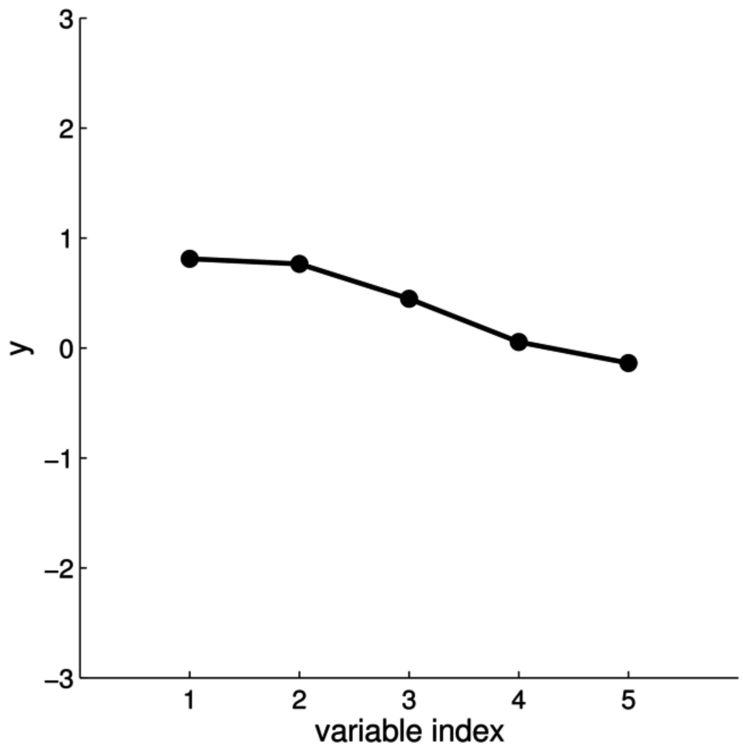




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

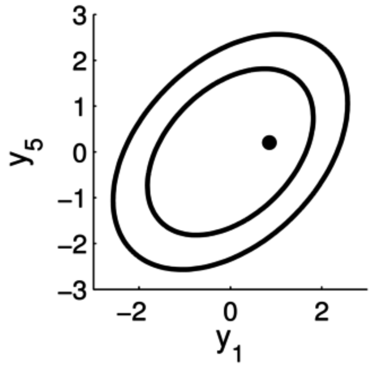
Visualizing Gaussian Processes

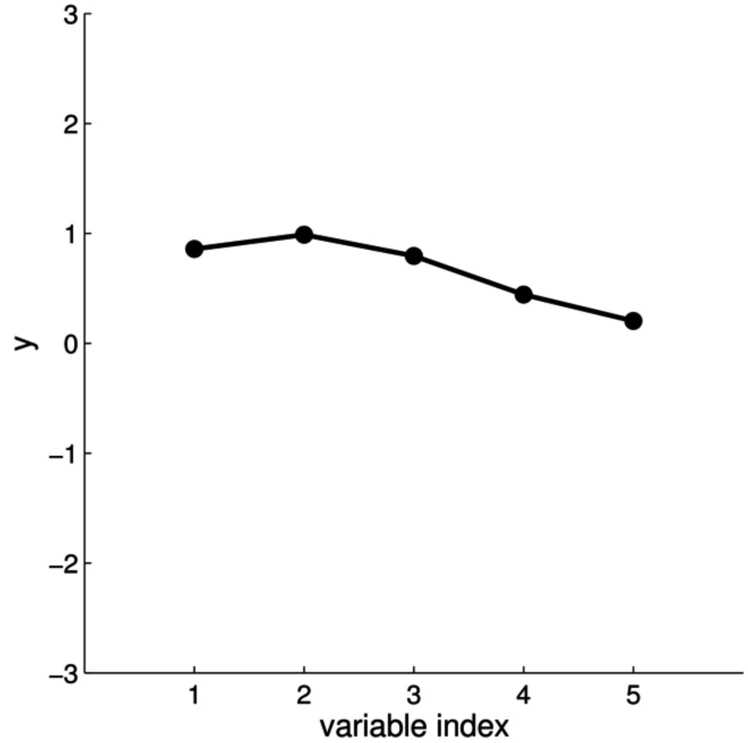




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

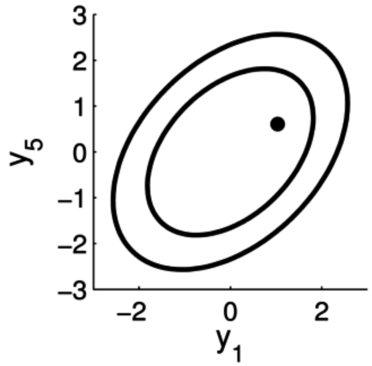
Visualizing Gaussian Processes

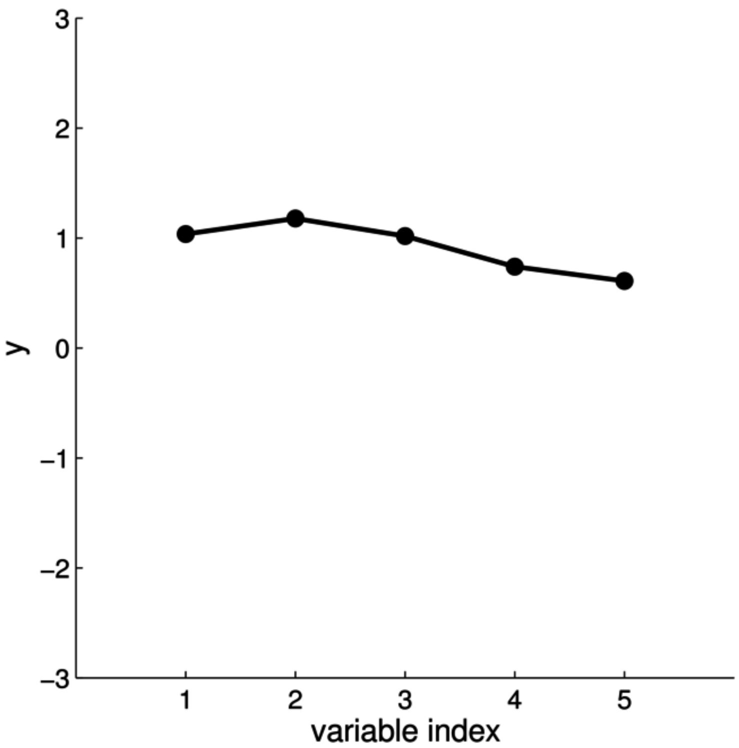




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

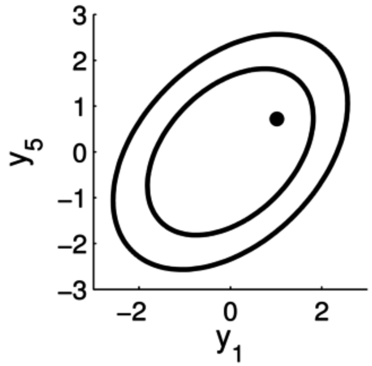
Visualizing Gaussian Processes

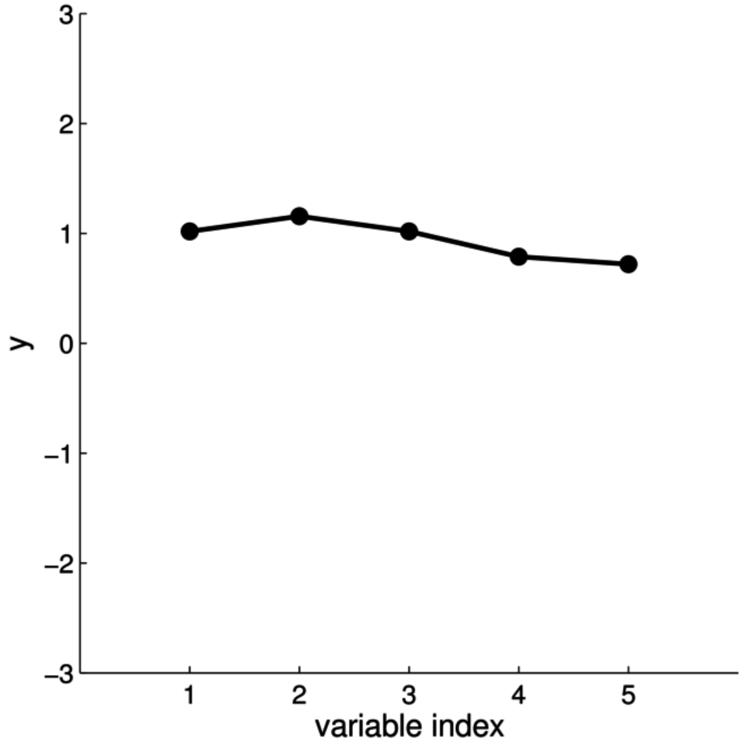




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

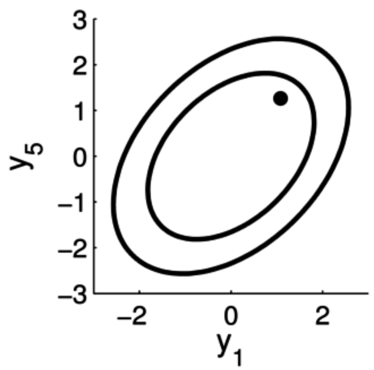
Visualizing Gaussian Processes

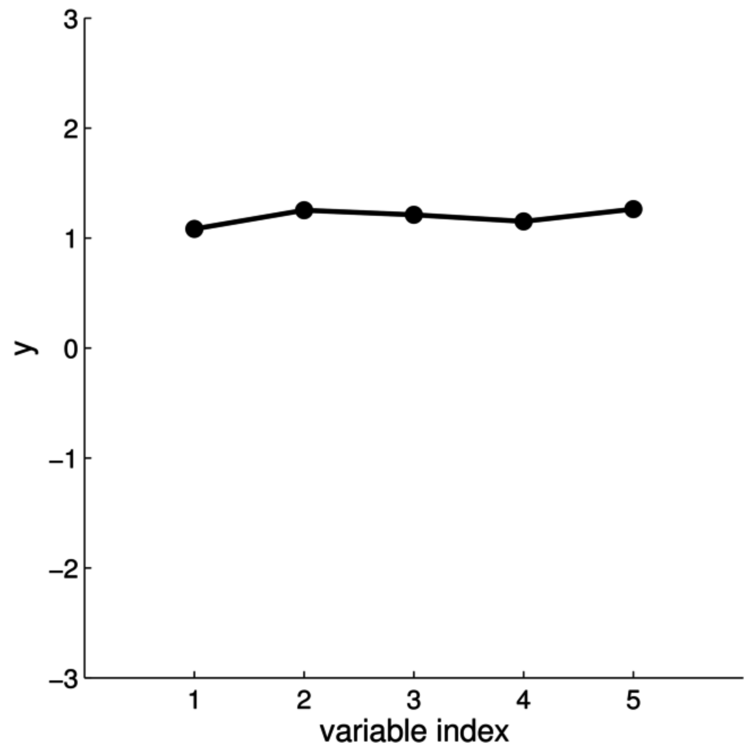




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

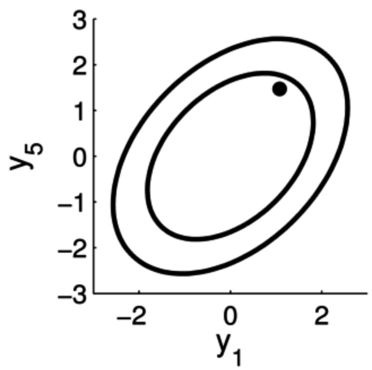
Visualizing Gaussian Processes



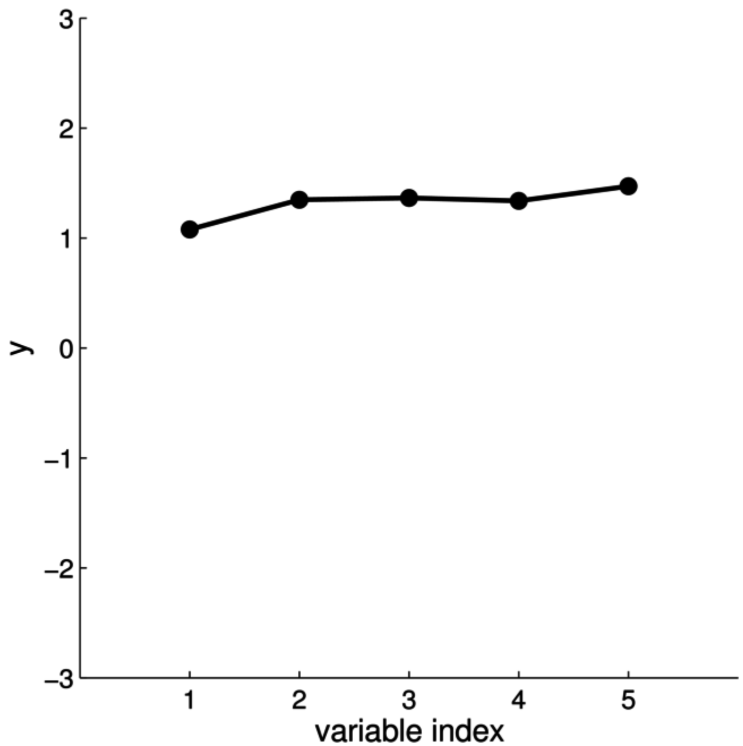


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

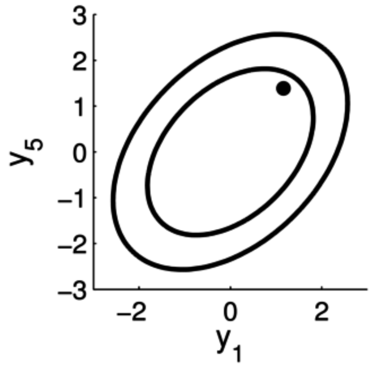
Visualizing Gaussian Processes

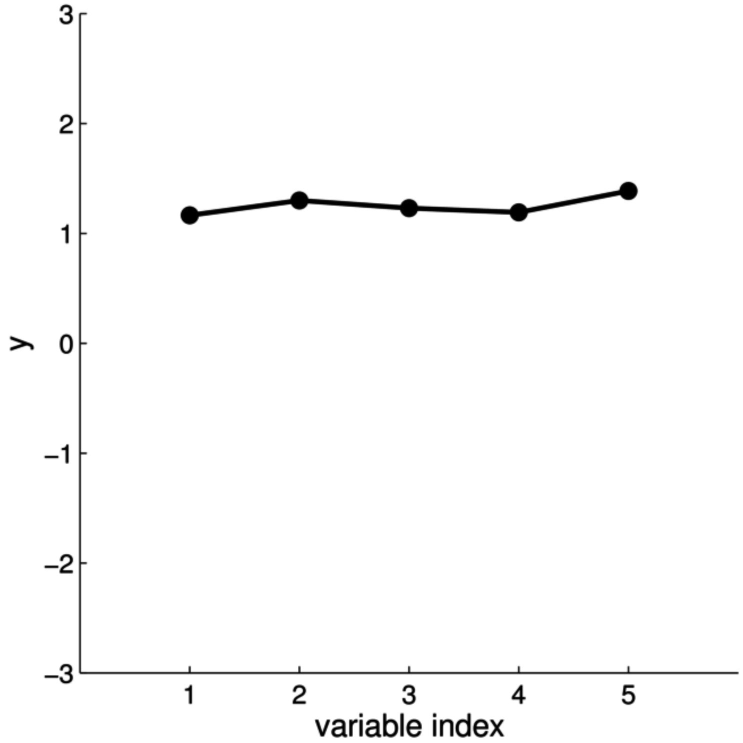


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



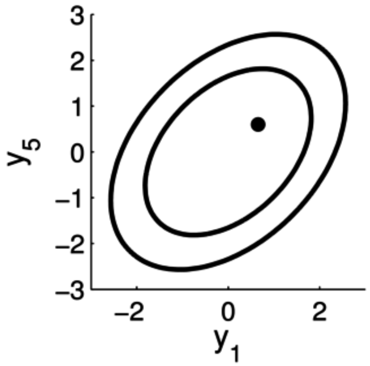
Visualizing Gaussian Processes

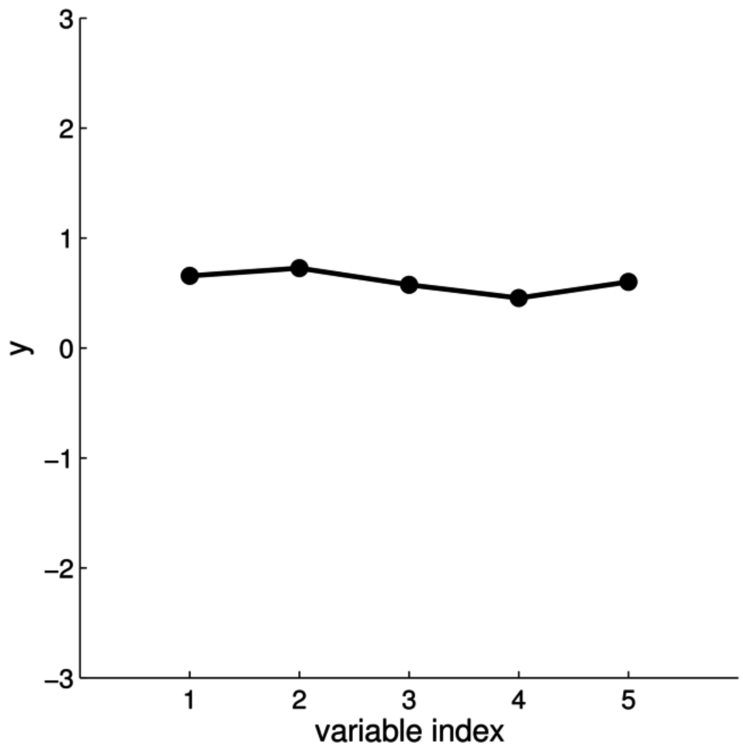




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

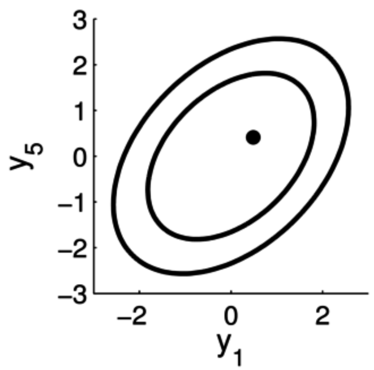
Visualizing Gaussian Processes

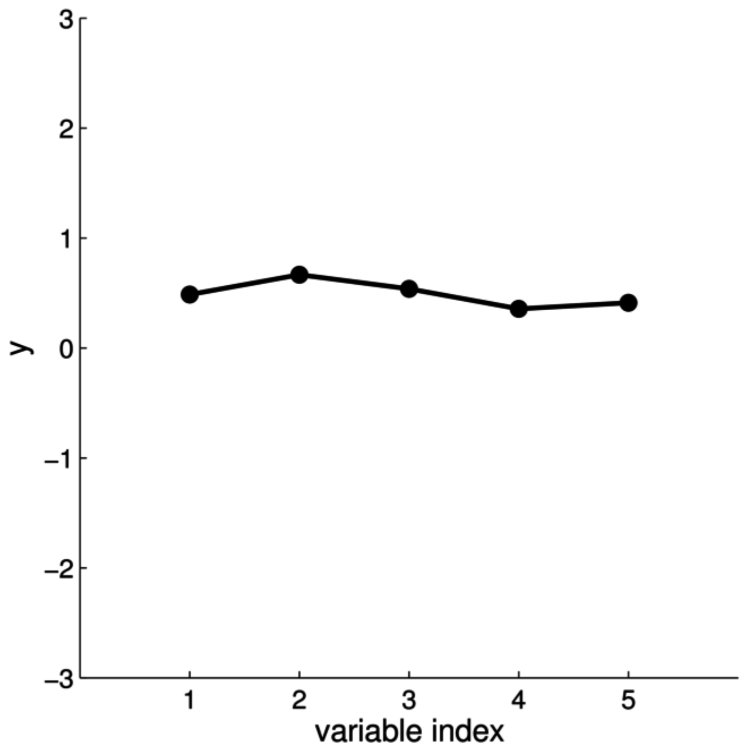




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

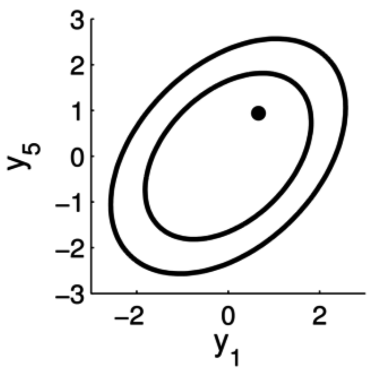
Visualizing Gaussian Processes

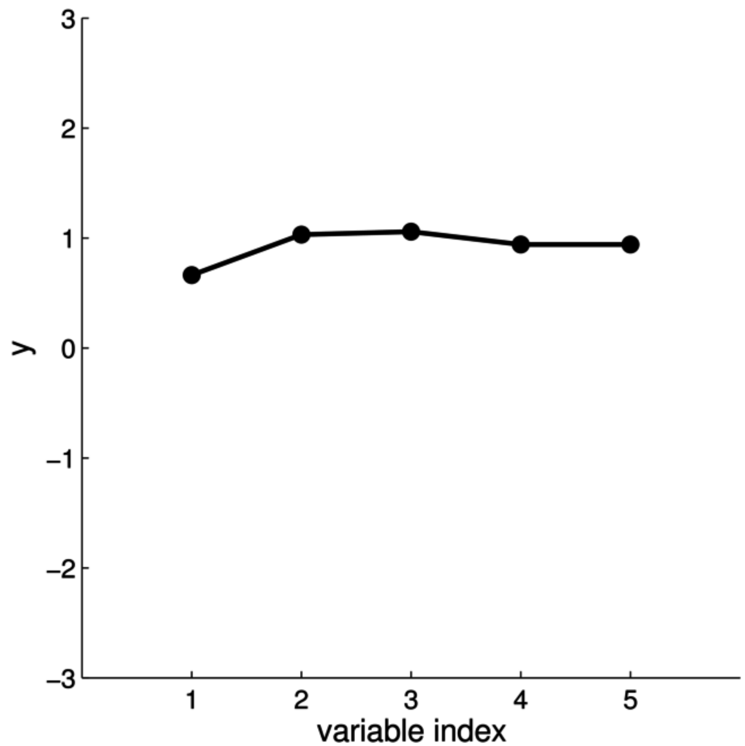




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

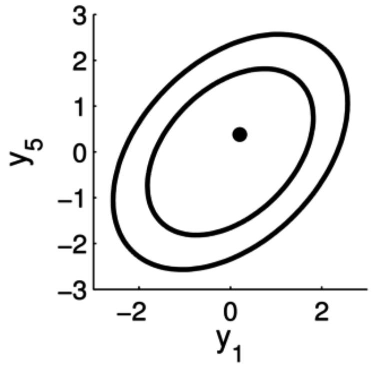
Visualizing Gaussian Processes

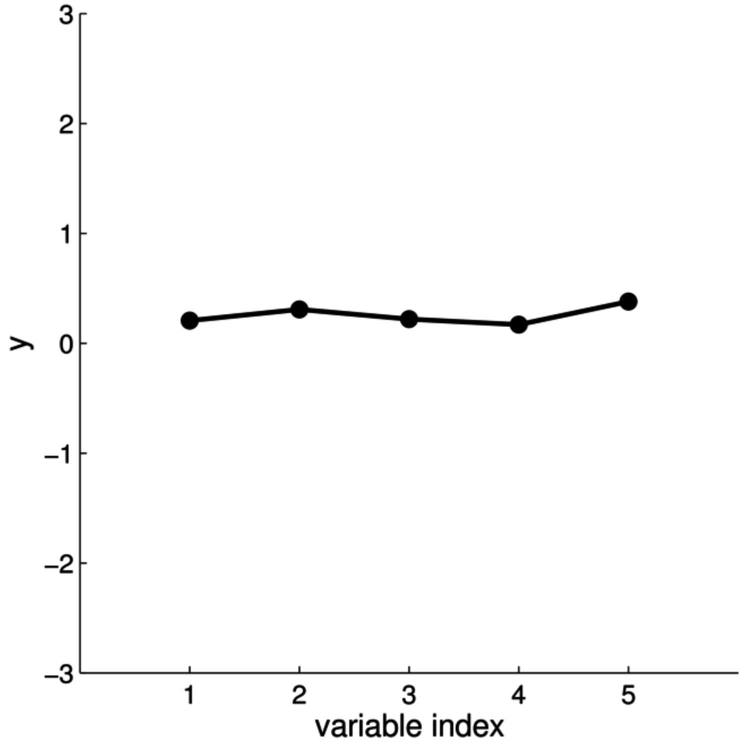




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

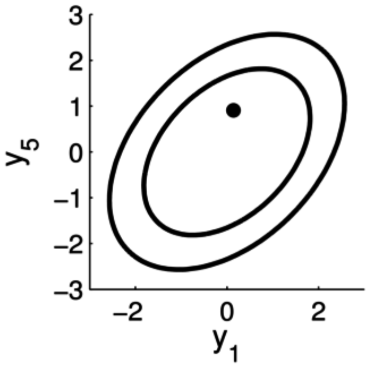
Visualizing Gaussian Processes



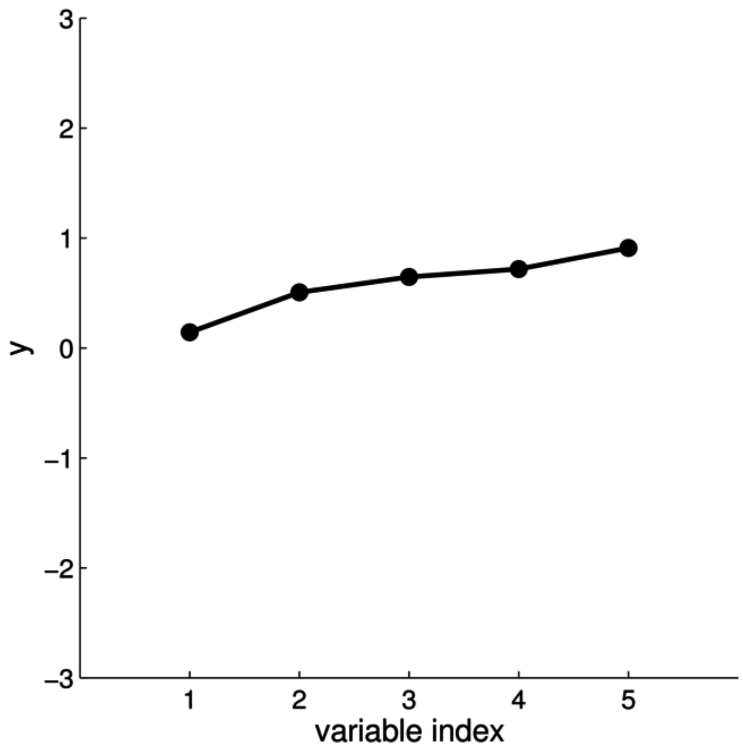


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

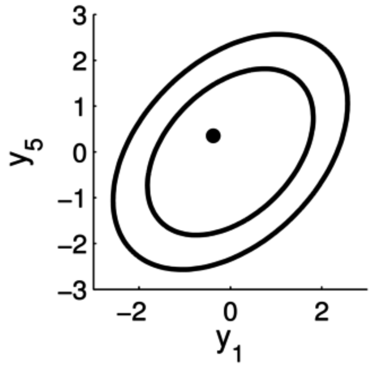
Visualizing Gaussian Processes

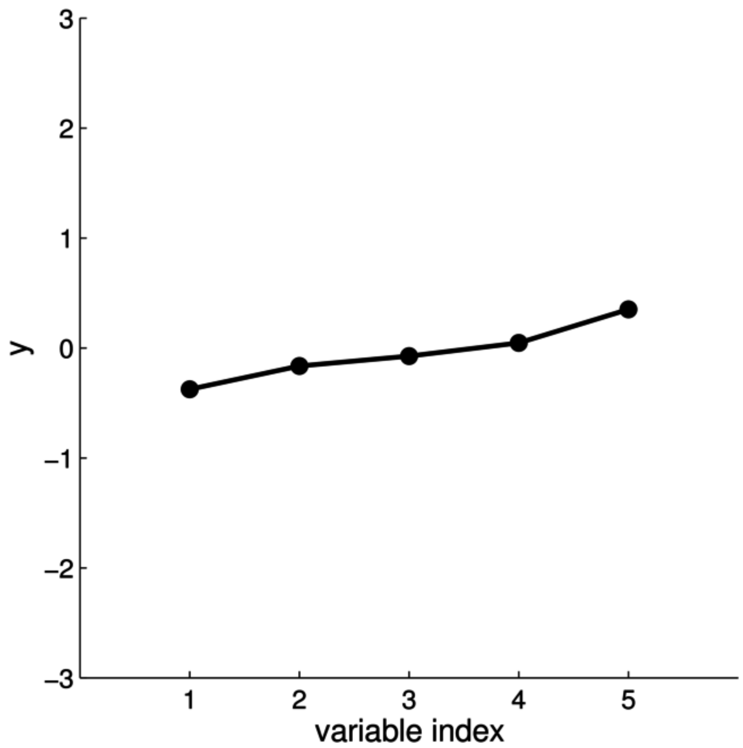


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



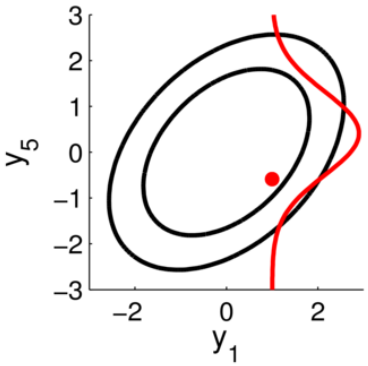
Visualizing Gaussian Processes

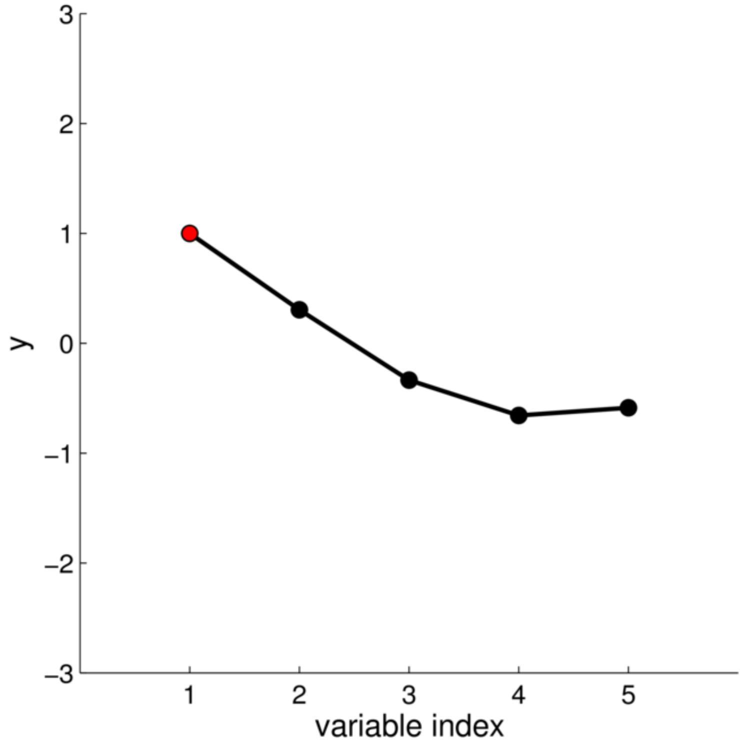




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

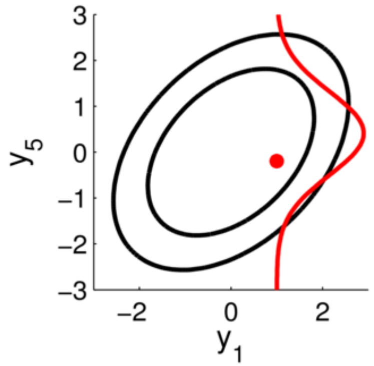
Visualizing Gaussian Processes

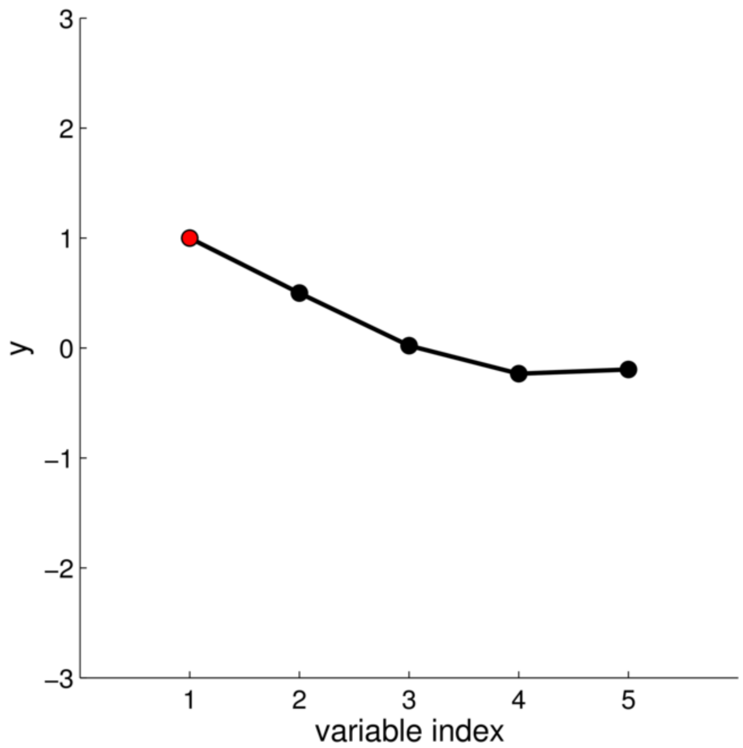




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

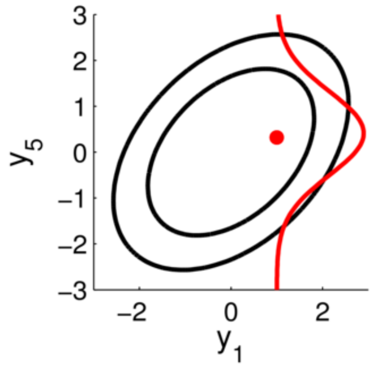
Visualizing Gaussian Processes

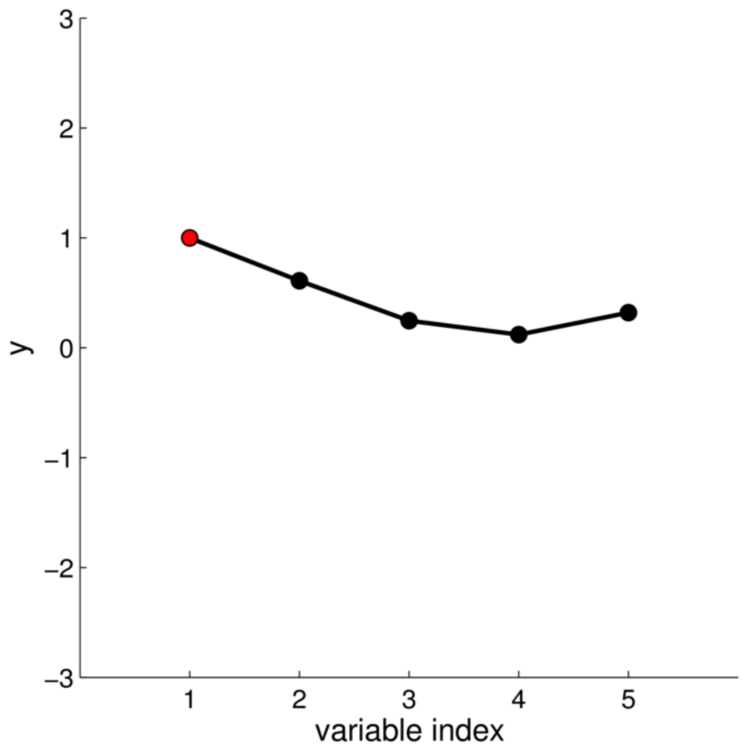




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

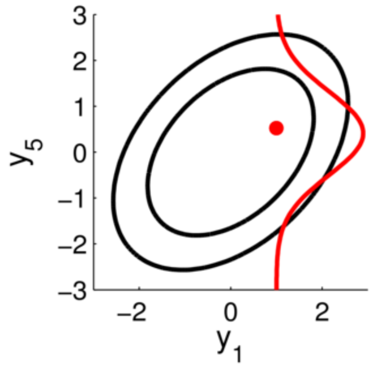
Visualizing Gaussian Processes

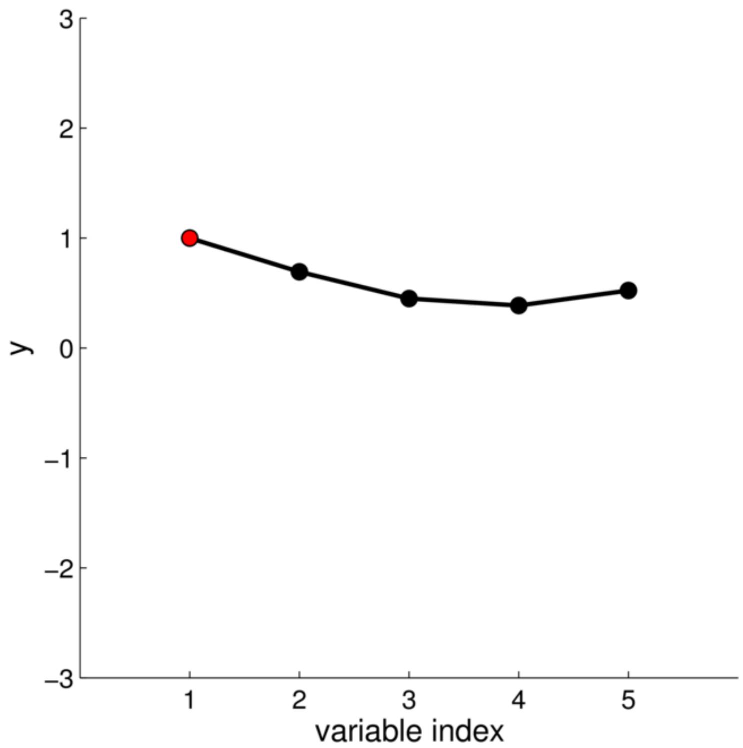




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

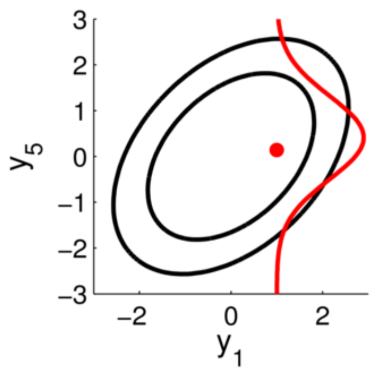
Visualizing Gaussian Processes



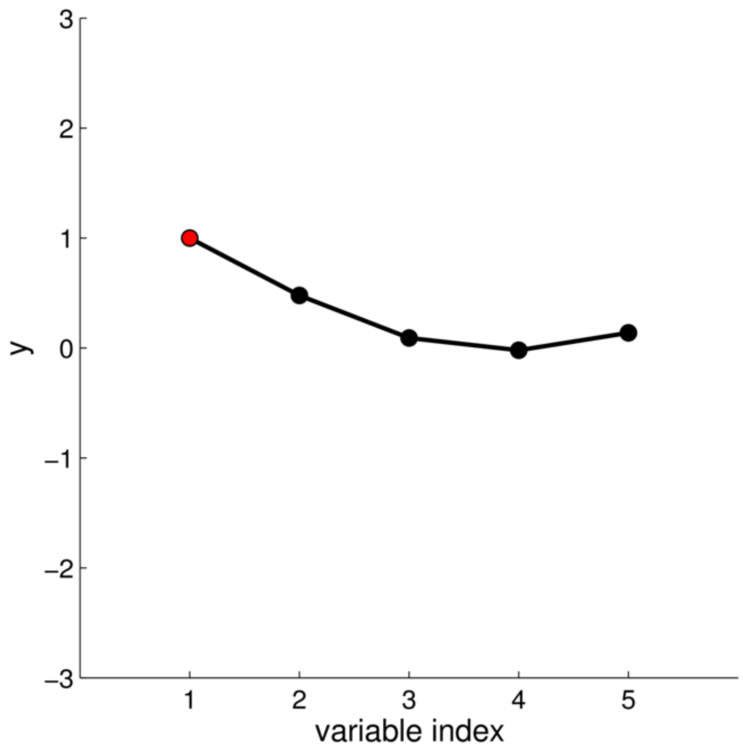


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

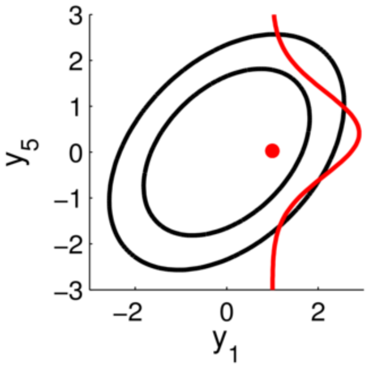
Visualizing Gaussian Processes

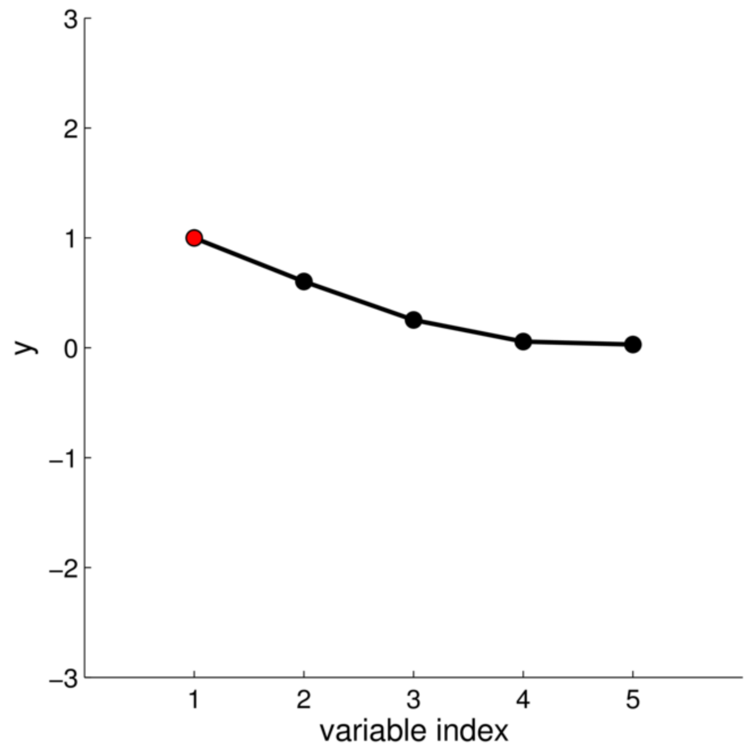


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



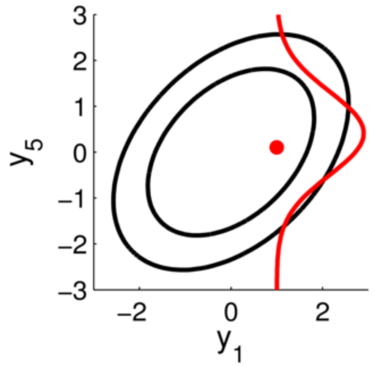
Visualizing Gaussian Processes

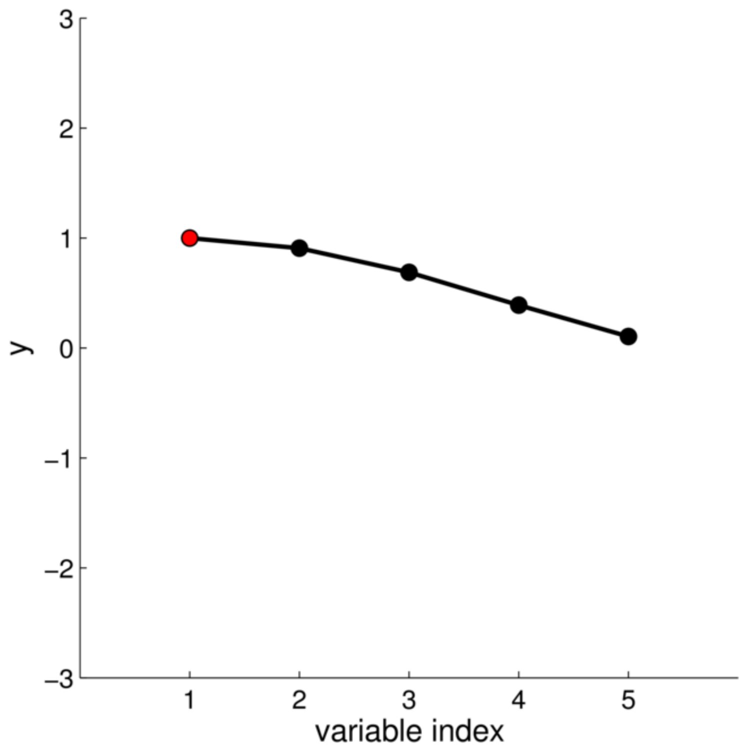




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

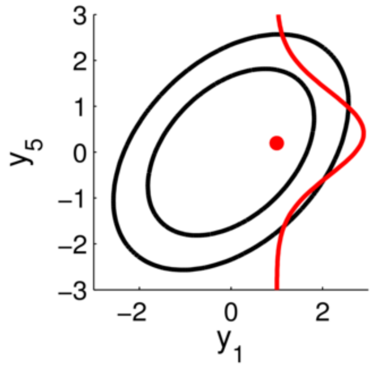
Visualizing Gaussian Processes

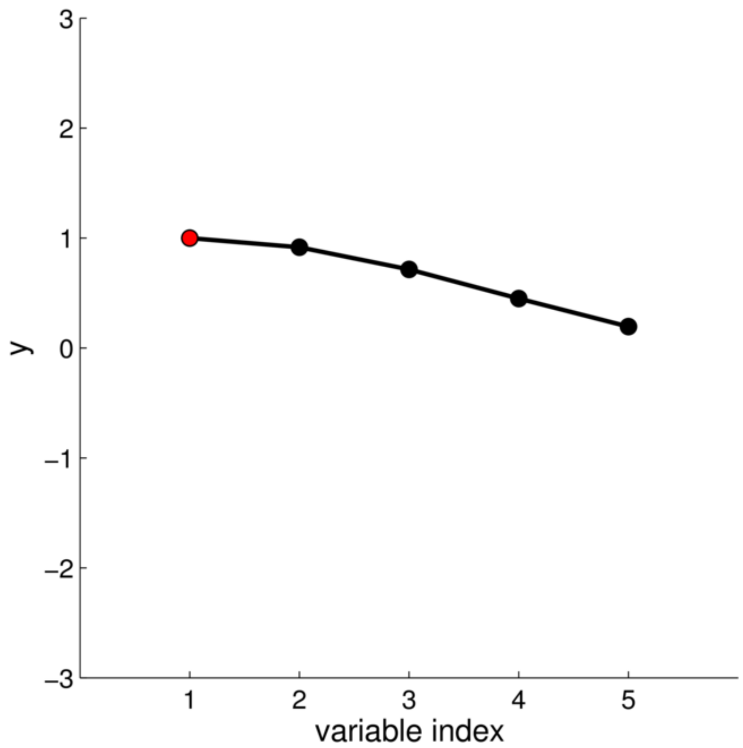




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

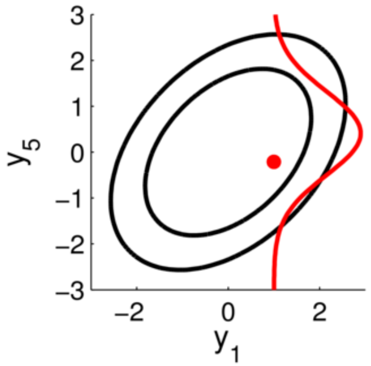
Visualizing Gaussian Processes

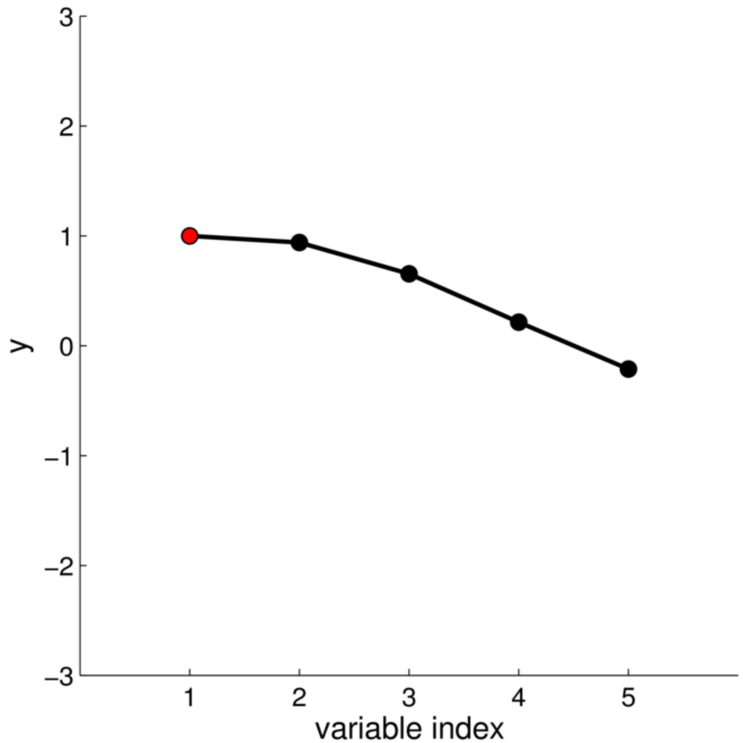




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

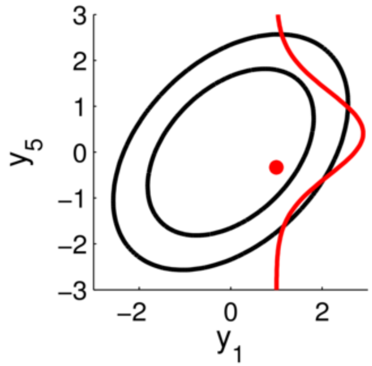
Visualizing Gaussian Processes

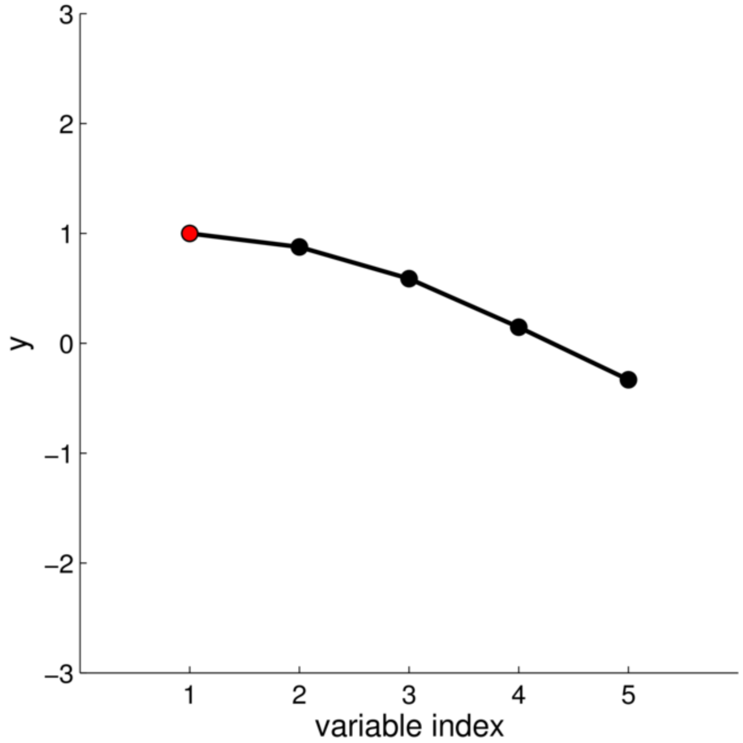




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

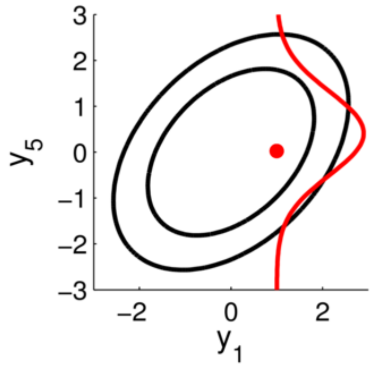
Visualizing Gaussian Processes



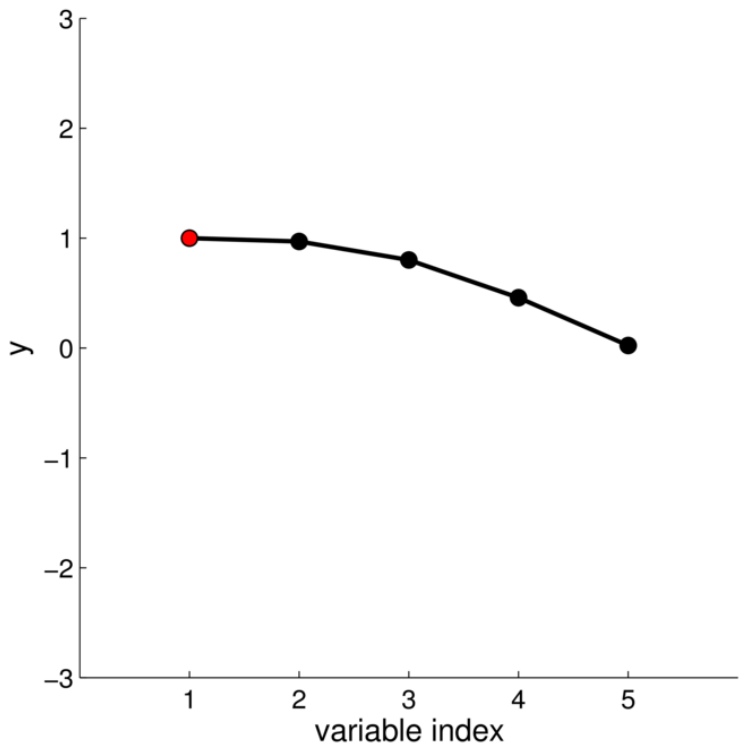


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

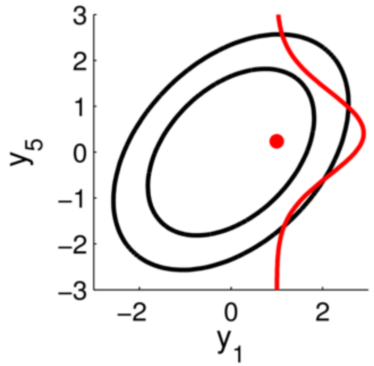
Visualizing Gaussian Processes

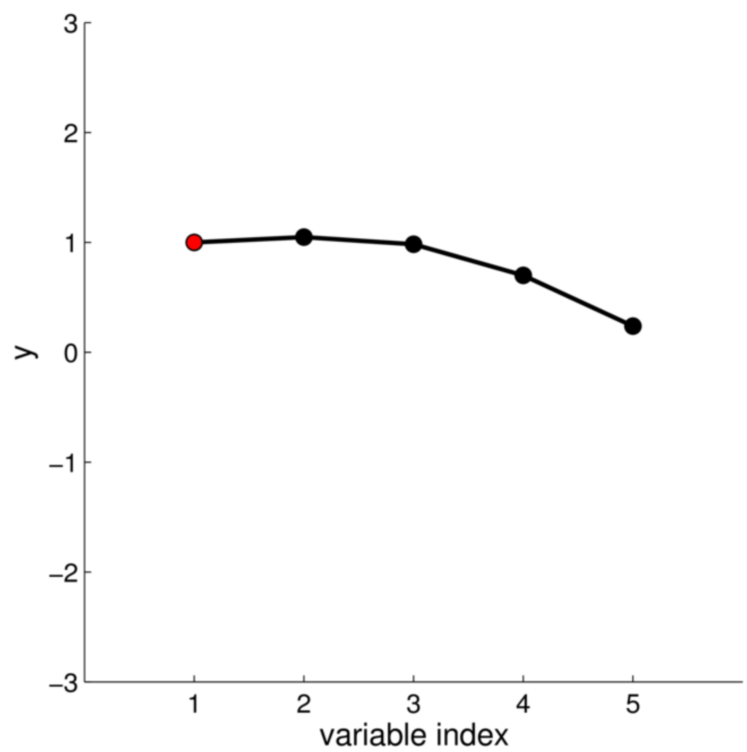


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



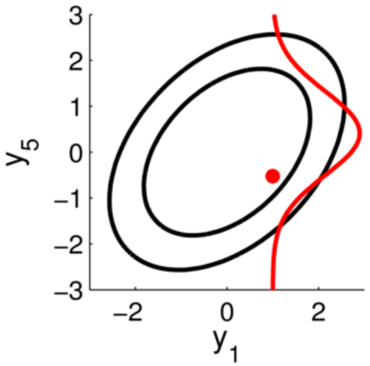
Visualizing Gaussian Processes

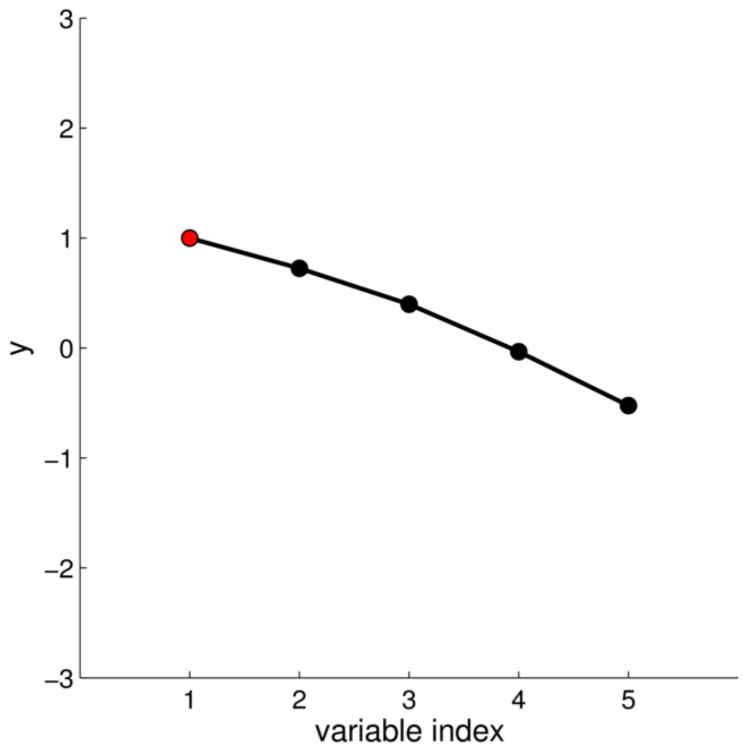




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

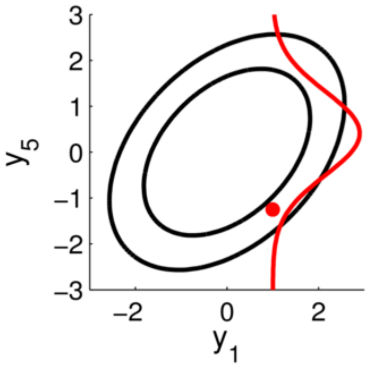
Visualizing Gaussian Processes

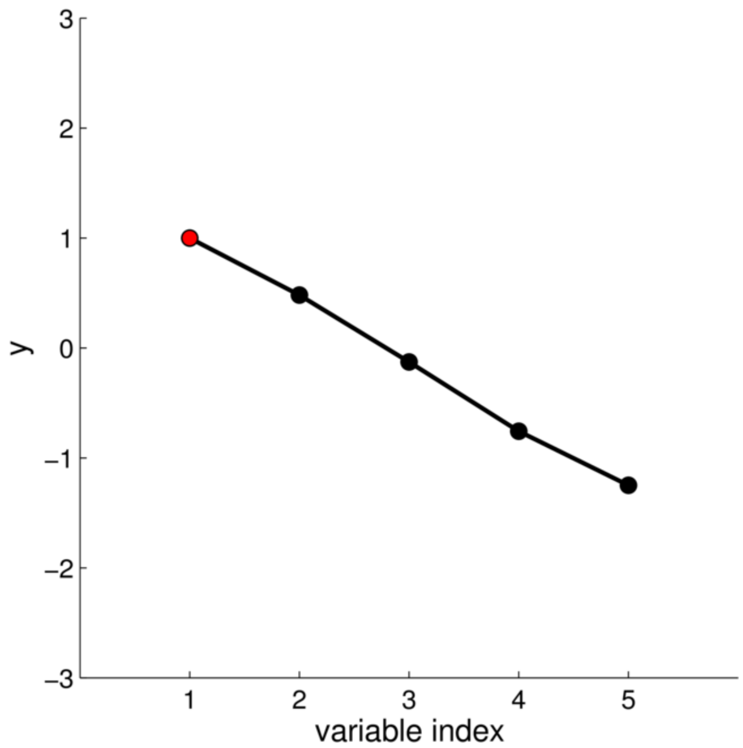




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

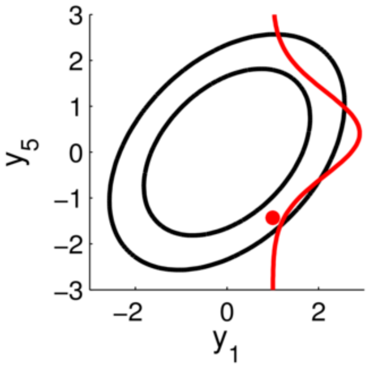
Visualizing Gaussian Processes



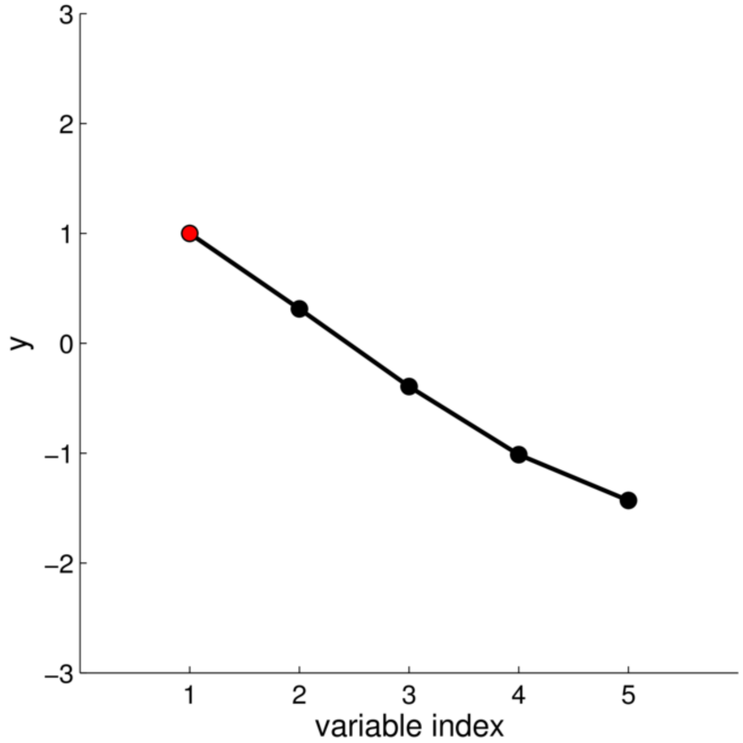


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

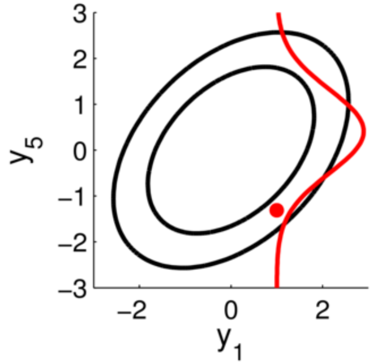
Visualizing Gaussian Processes

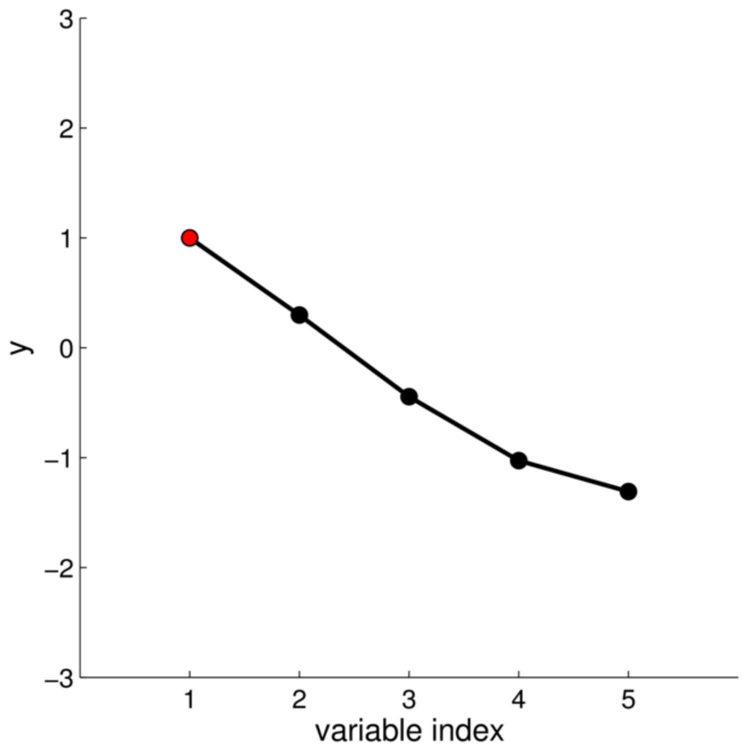


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



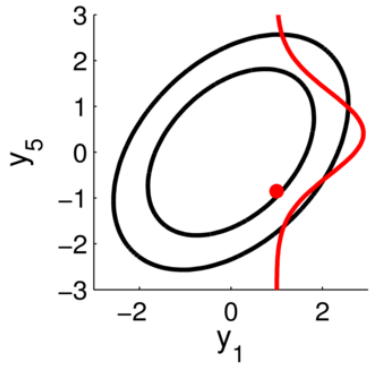
Visualizing Gaussian Processes

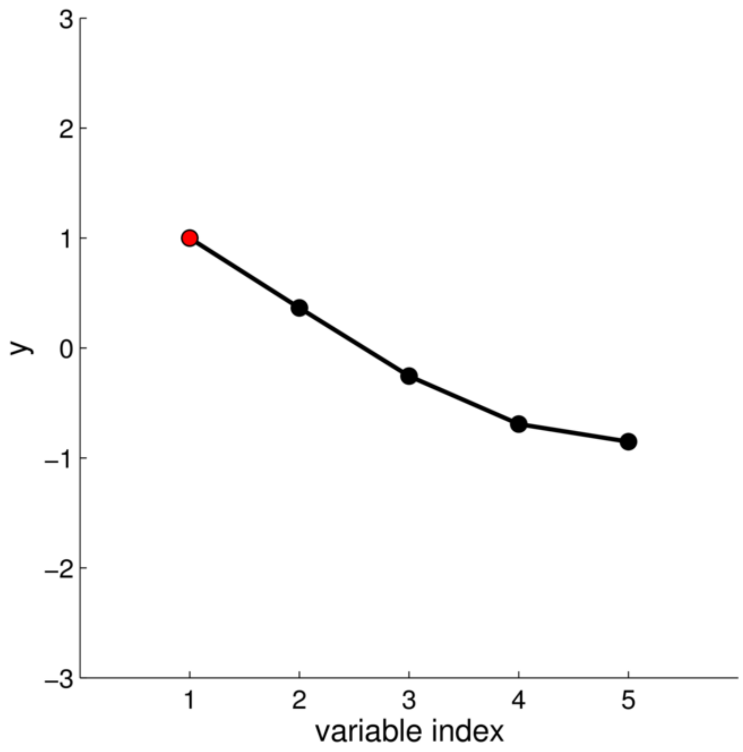




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

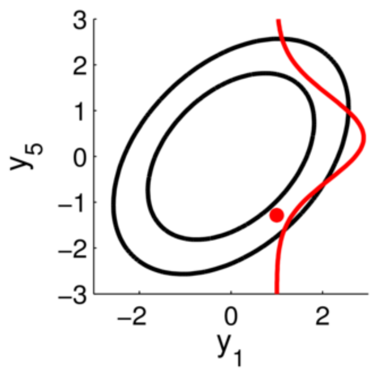
Visualizing Gaussian Processes

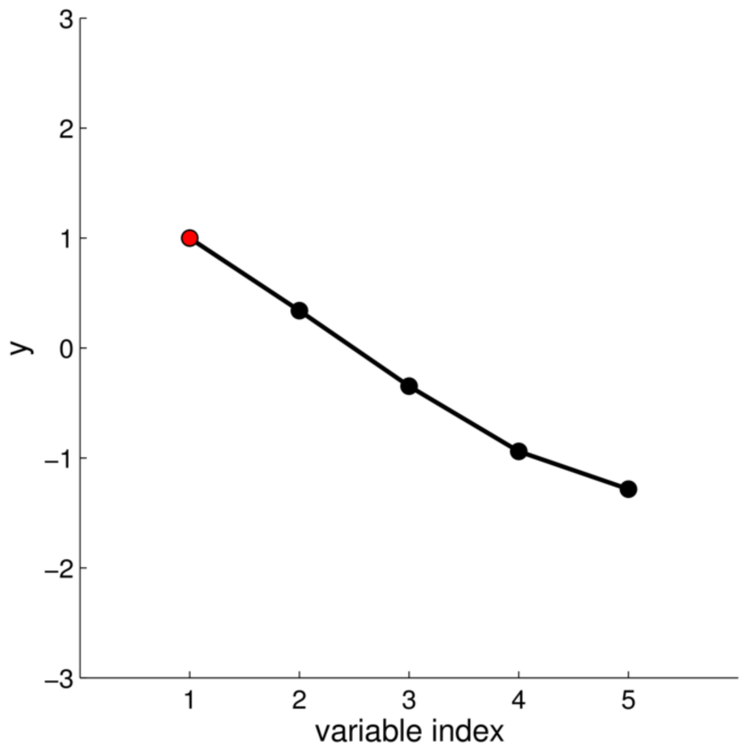




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

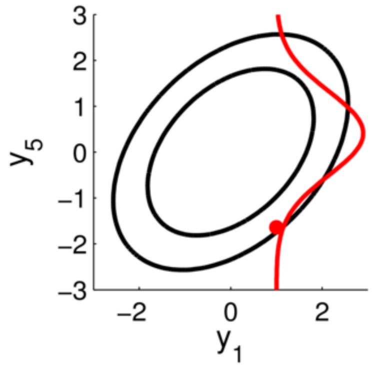
Visualizing Gaussian Processes

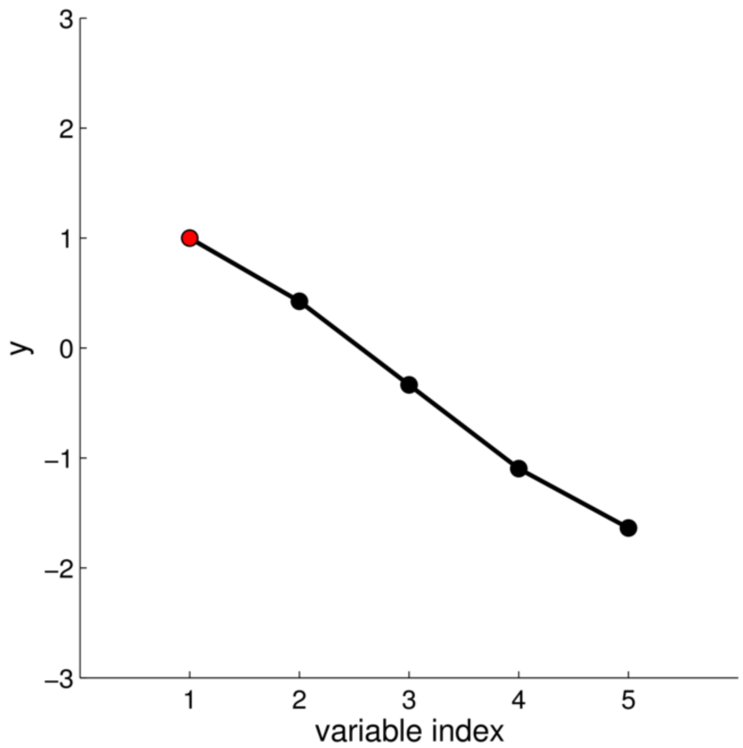




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

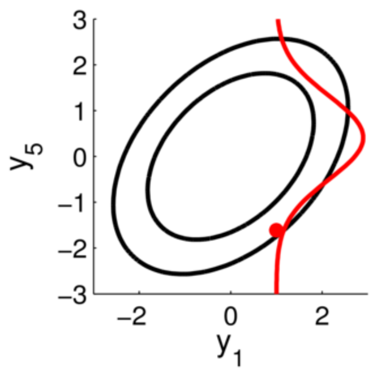
Visualizing Gaussian Processes

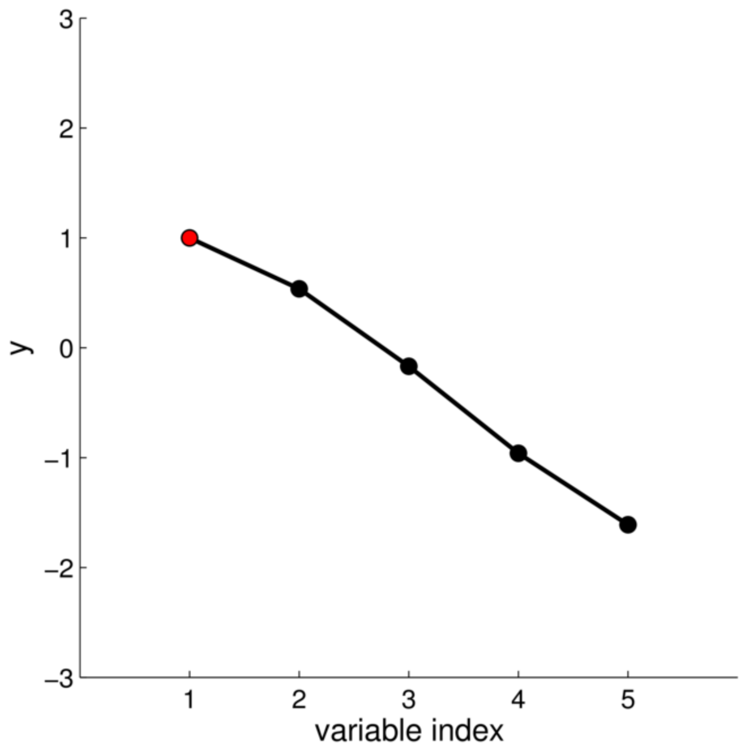




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

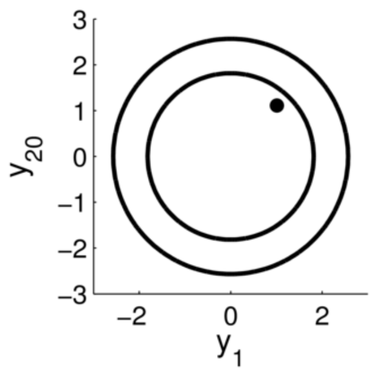
Visualizing Gaussian Processes



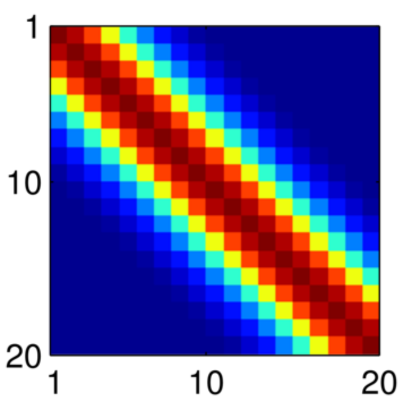


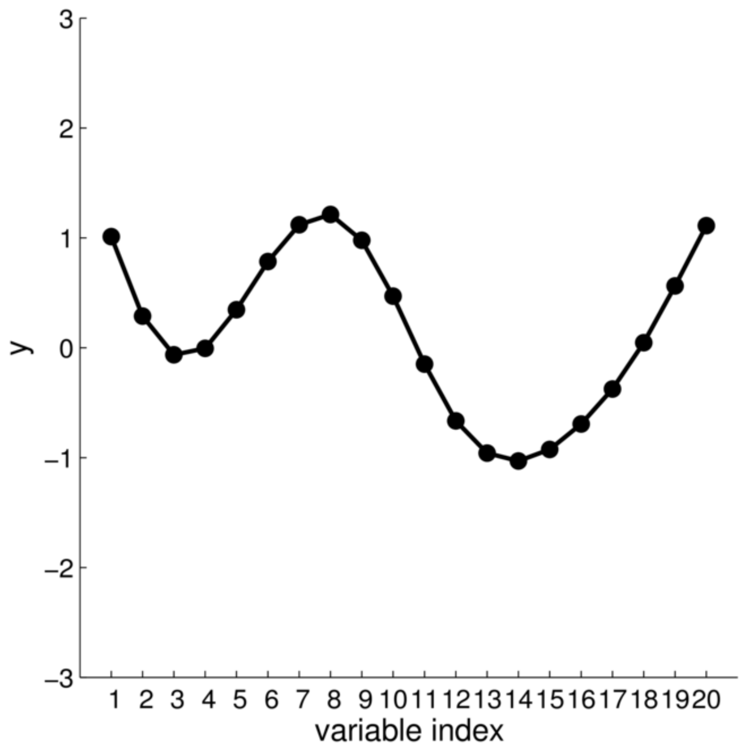
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

Visualizing Gaussian Processes

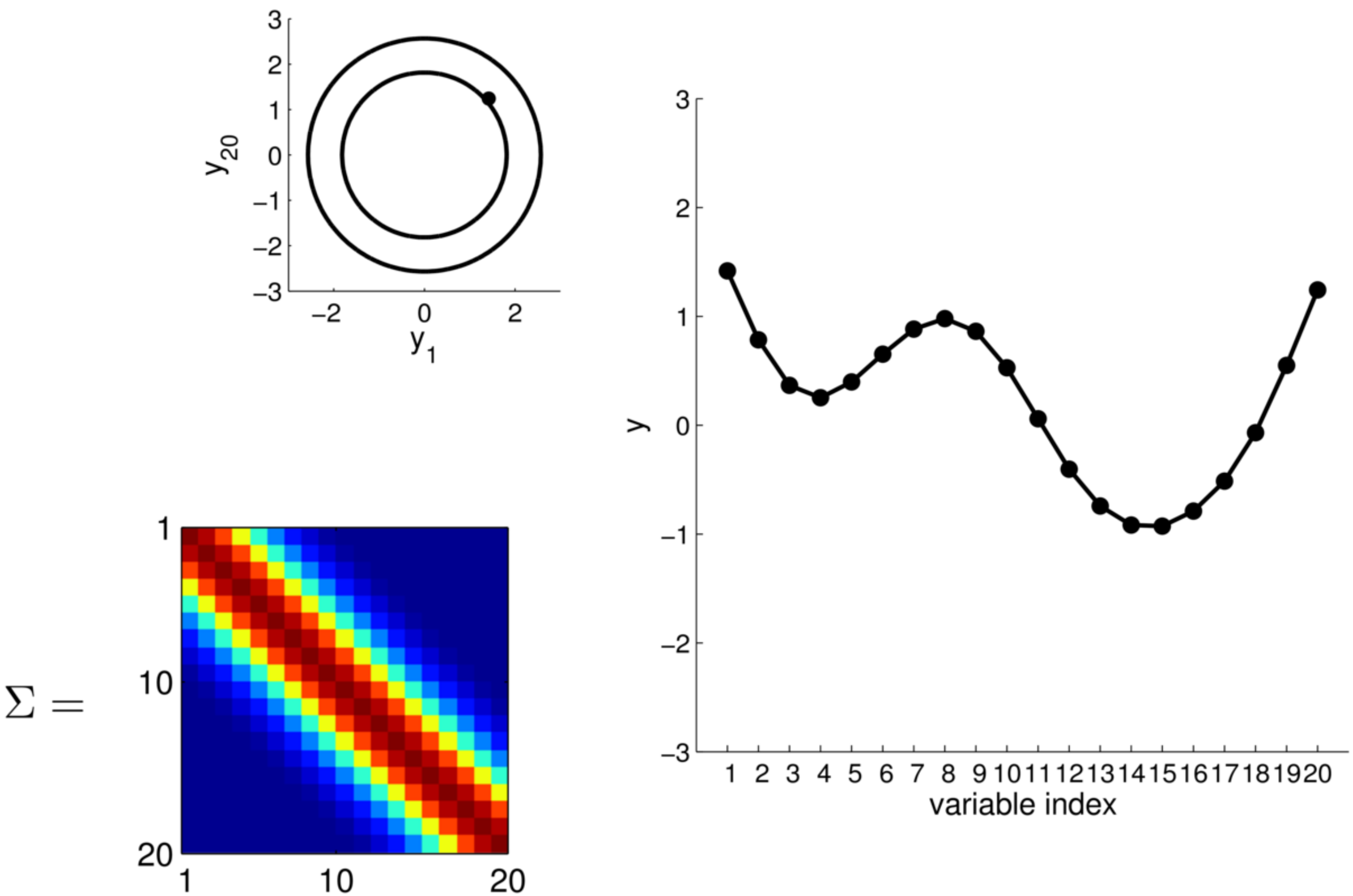


$\Sigma =$

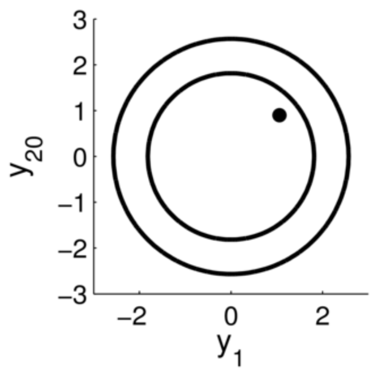


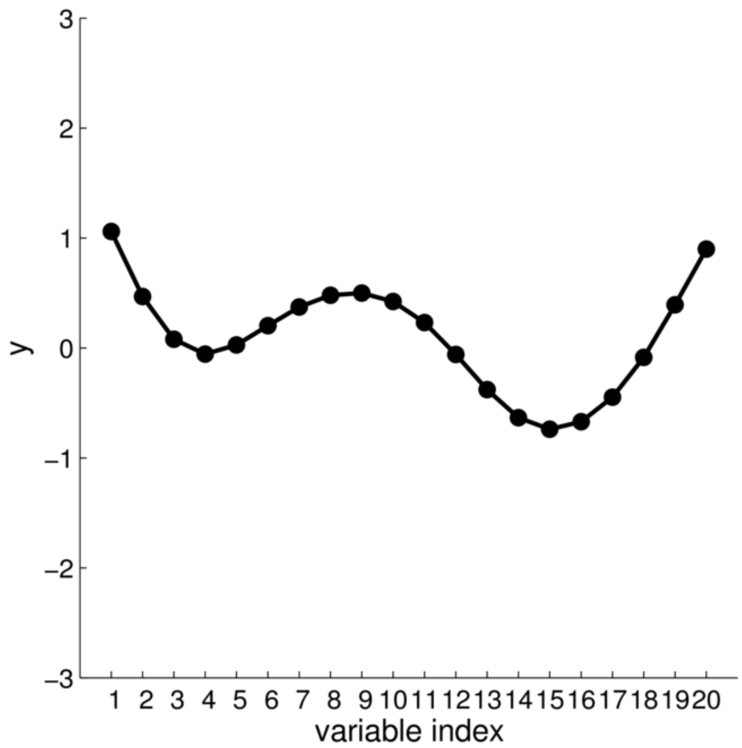


Visualizing Gaussian Processes

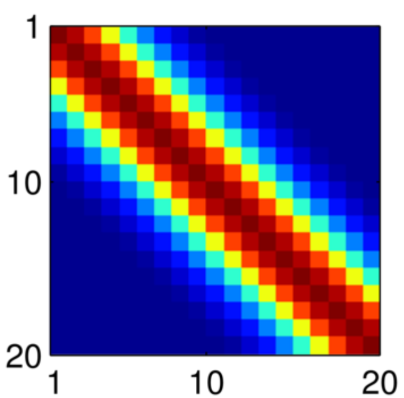


Visualizing Gaussian Processes

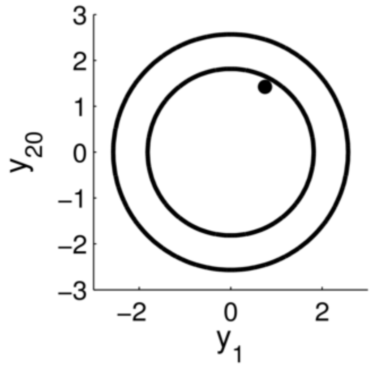


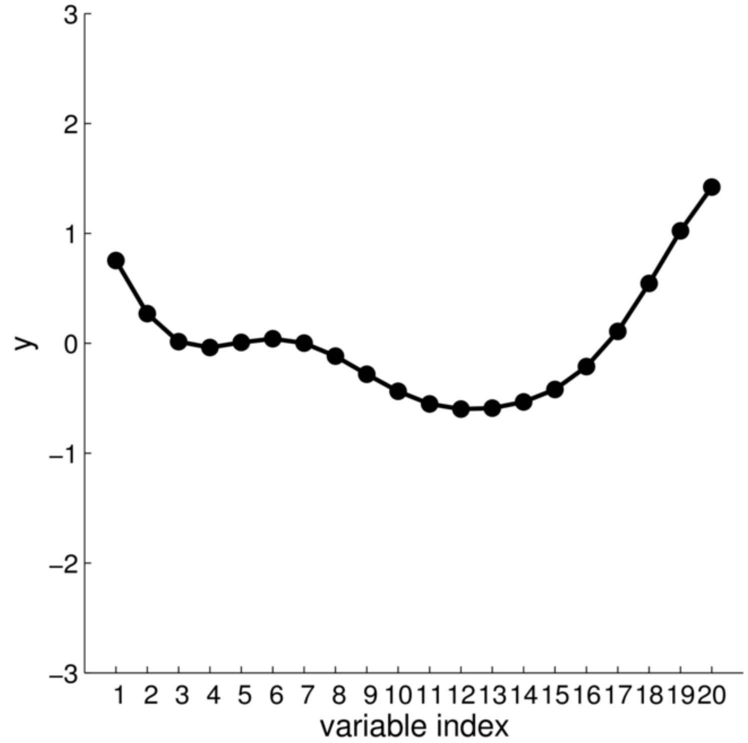


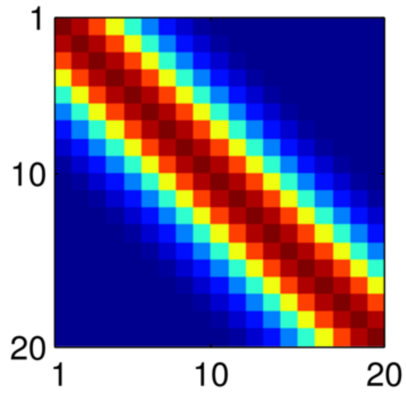
$\Sigma =$



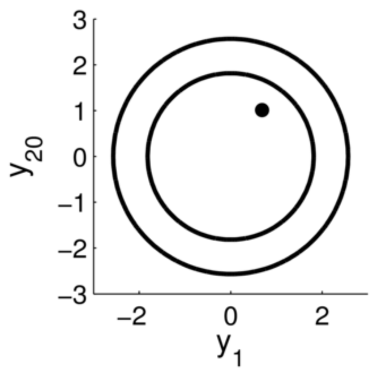
Visualizing Gaussian Processes

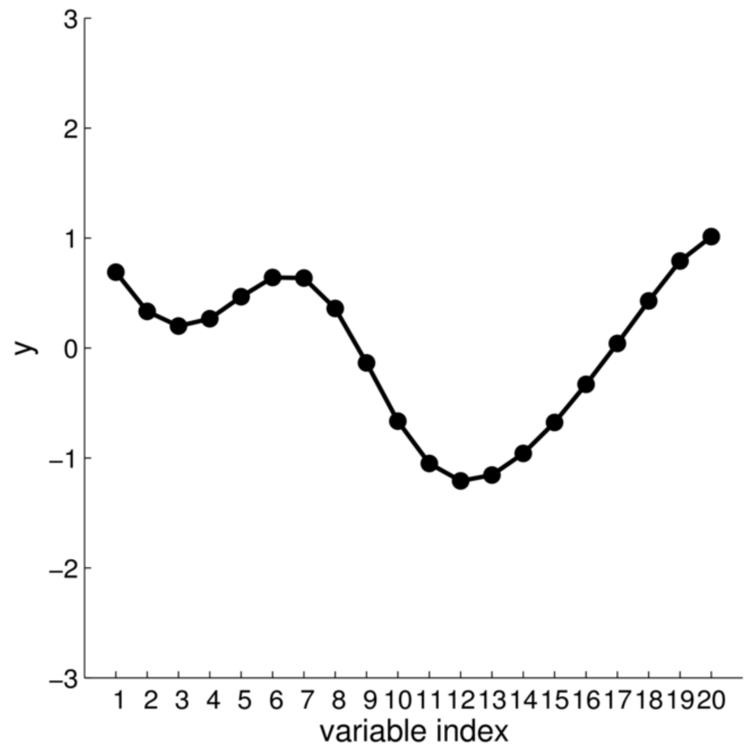




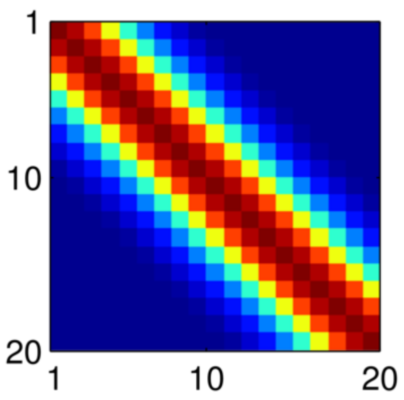
$\Sigma =$ 

Visualizing Gaussian Processes

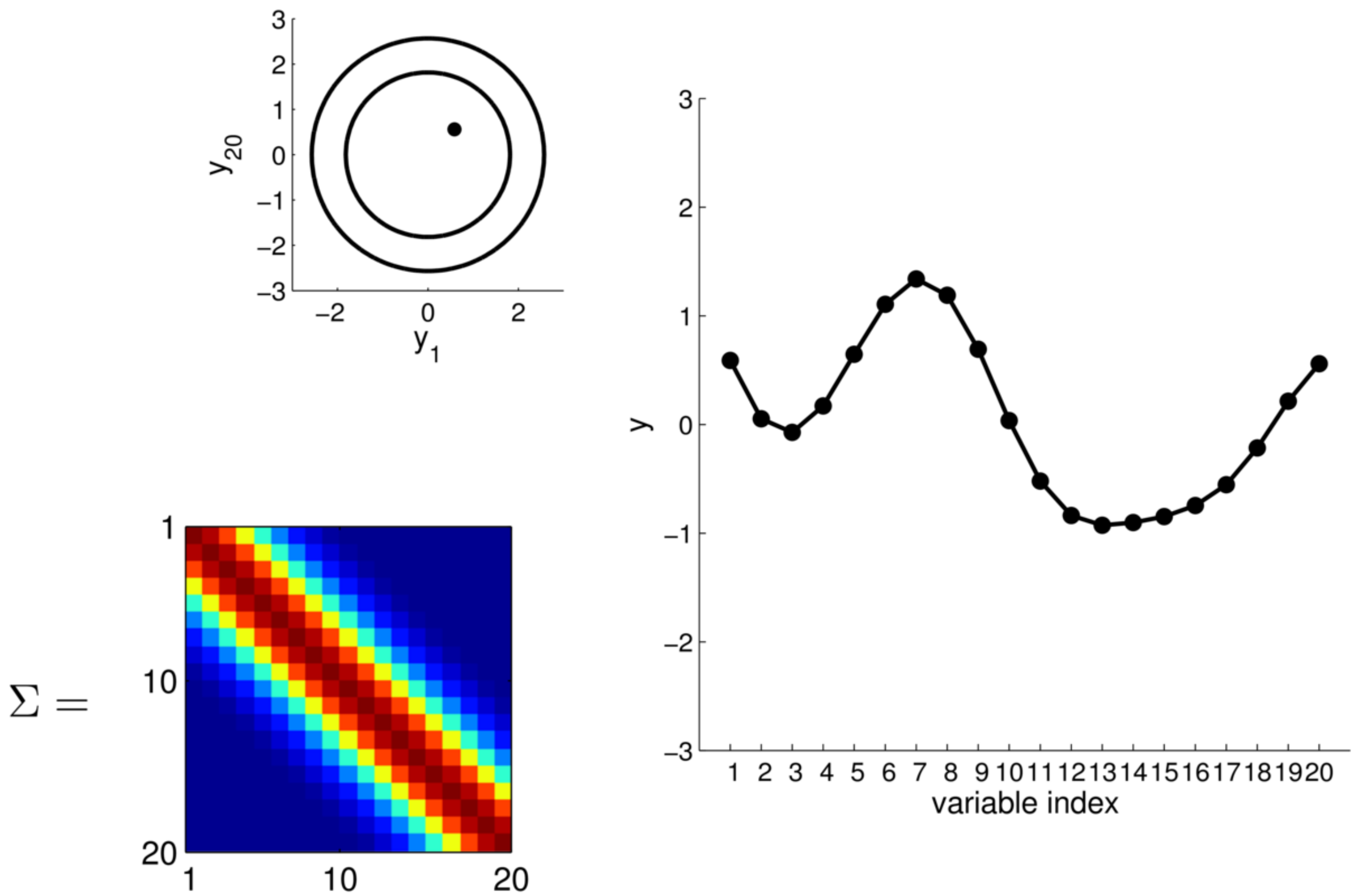




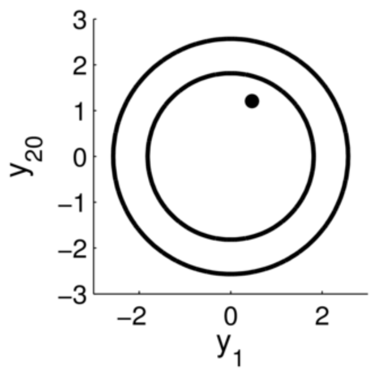
$\Sigma =$



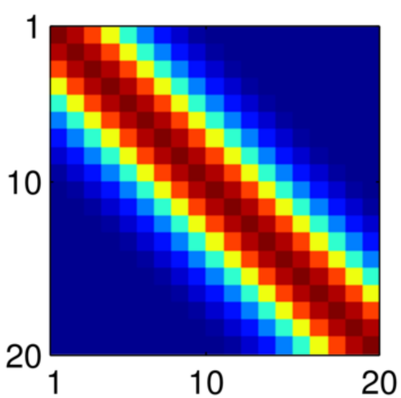
Visualizing Gaussian Processes

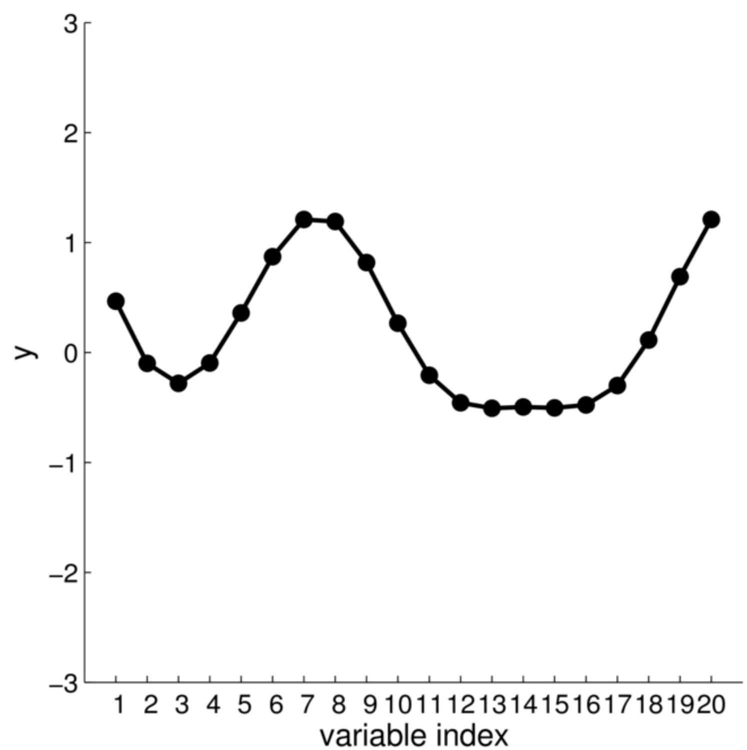


Visualizing Gaussian Processes

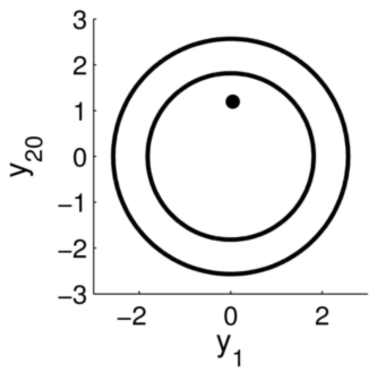


$\Sigma =$

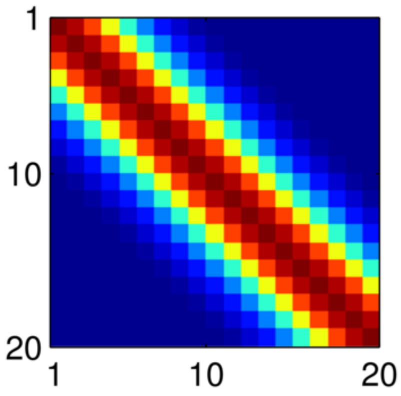


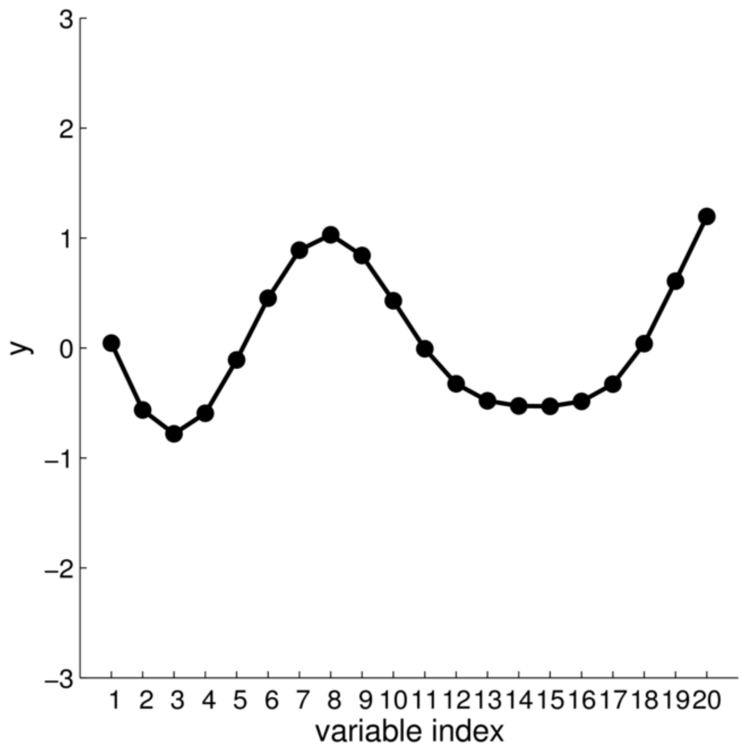


Visualizing Gaussian Processes

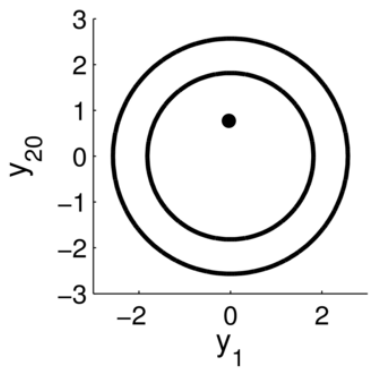


$\Sigma =$

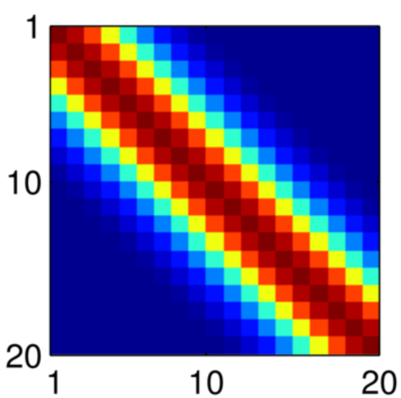


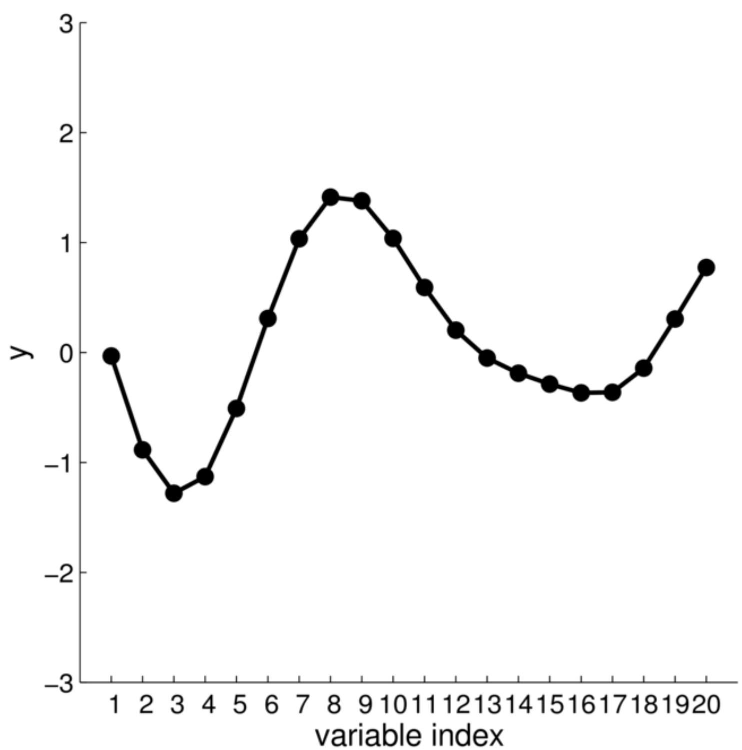


Visualizing Gaussian Processes

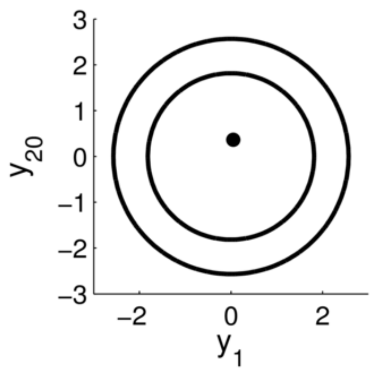


$\Sigma =$

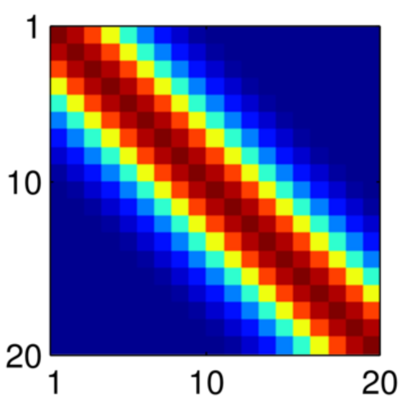


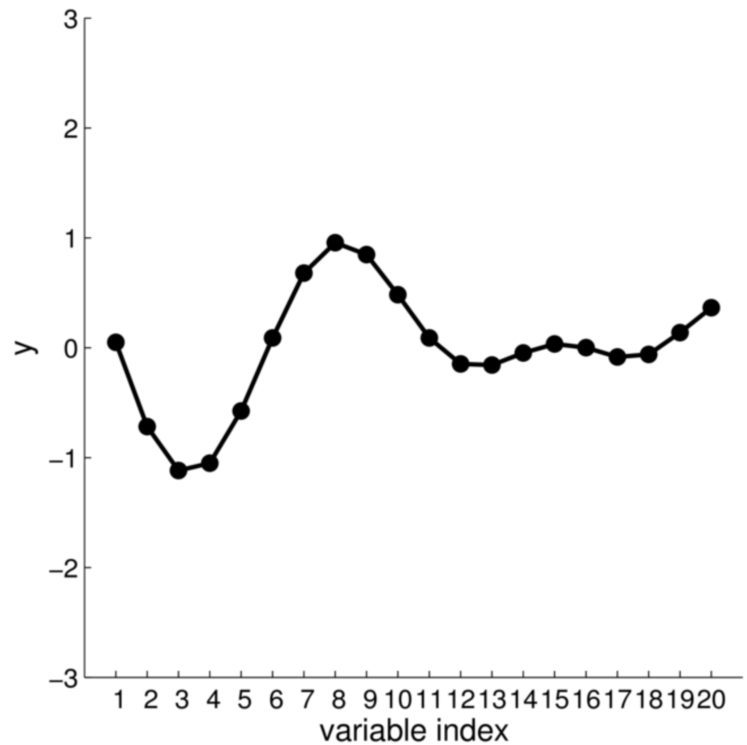


Visualizing Gaussian Processes

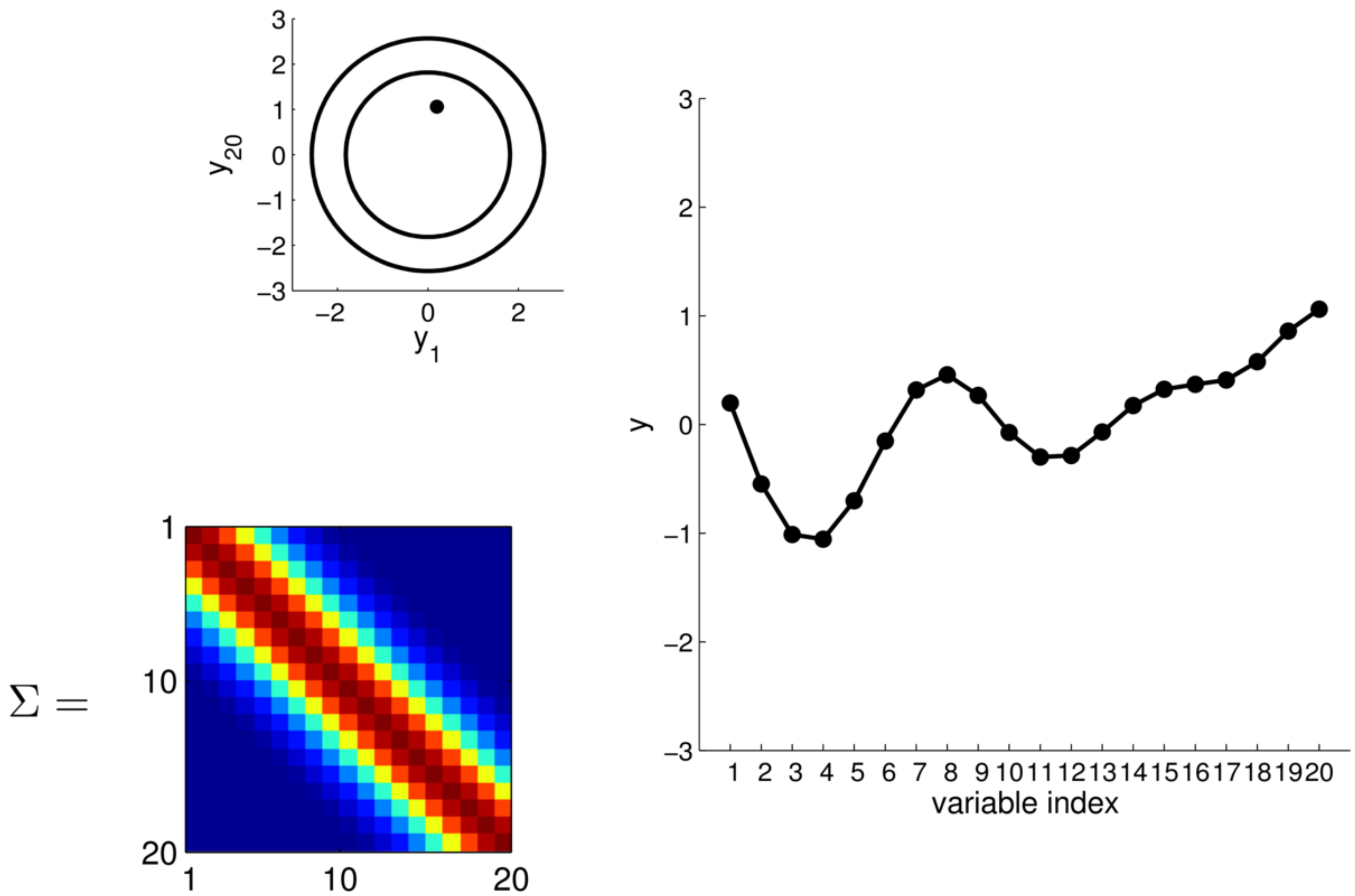


$\Sigma =$

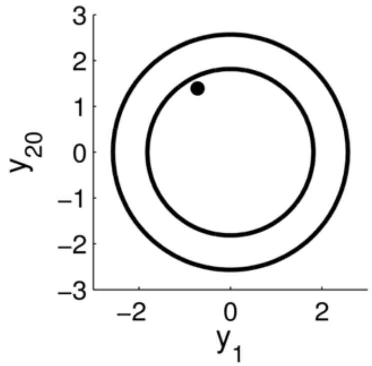


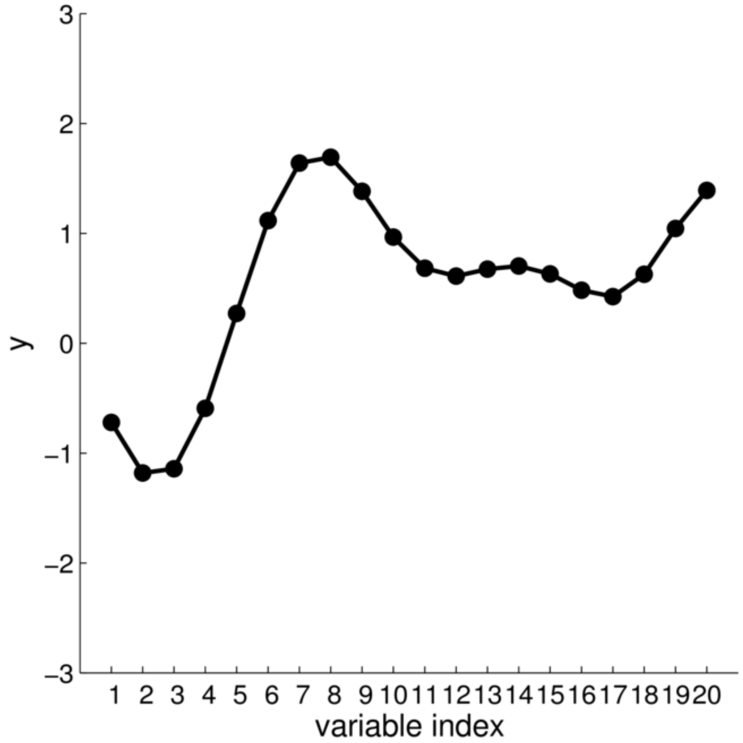


Visualizing Gaussian Processes

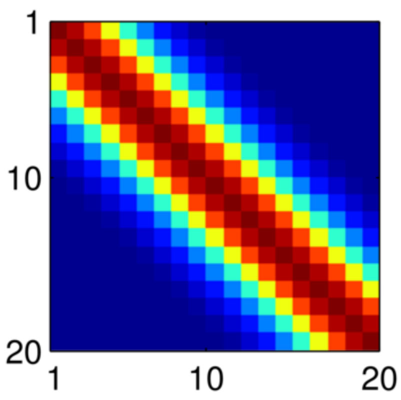


Visualizing Gaussian Processes

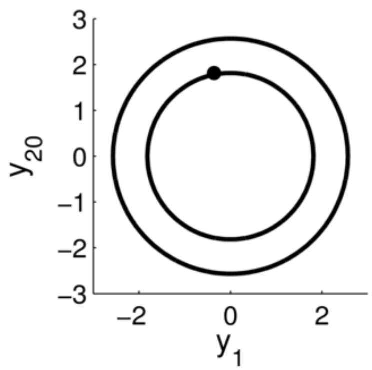


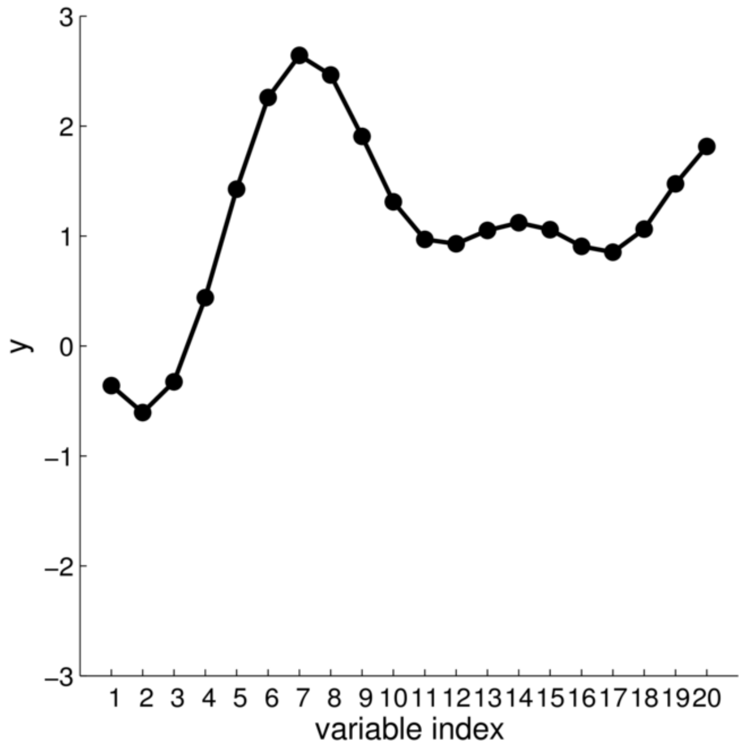


$\Sigma =$

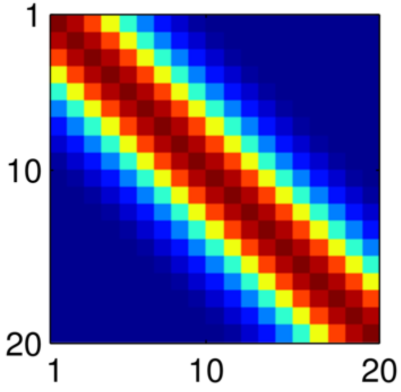


Visualizing Gaussian Processes

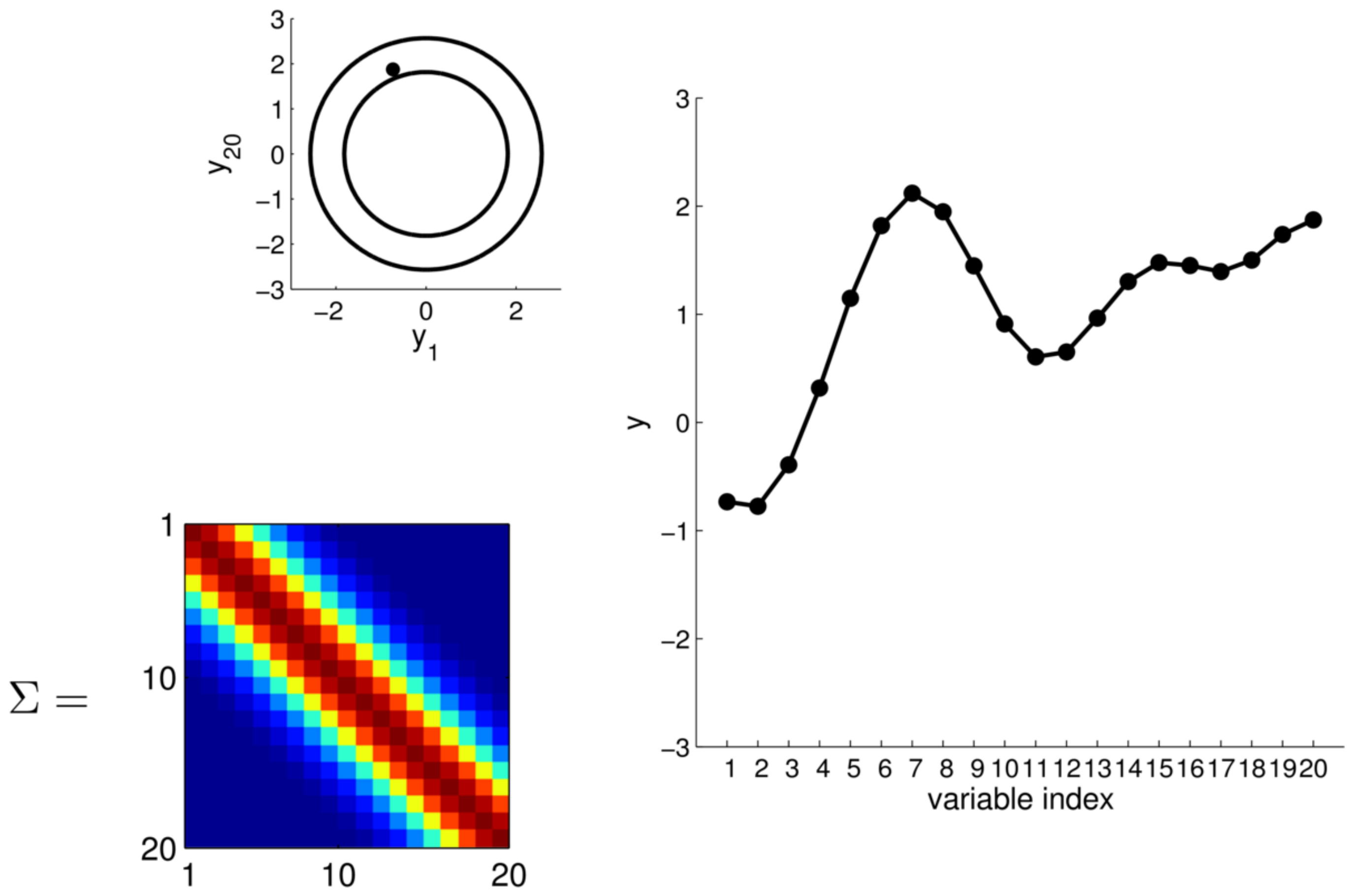




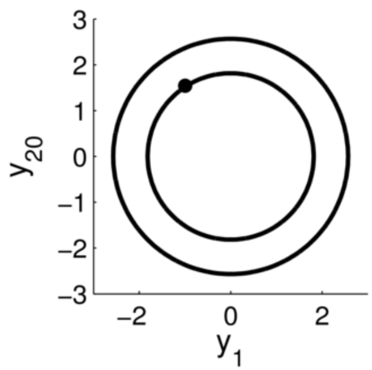
$\Sigma =$



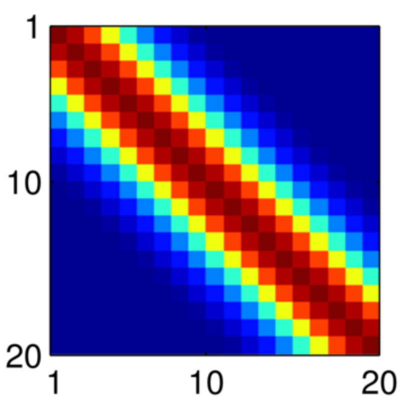
Visualizing Gaussian Processes

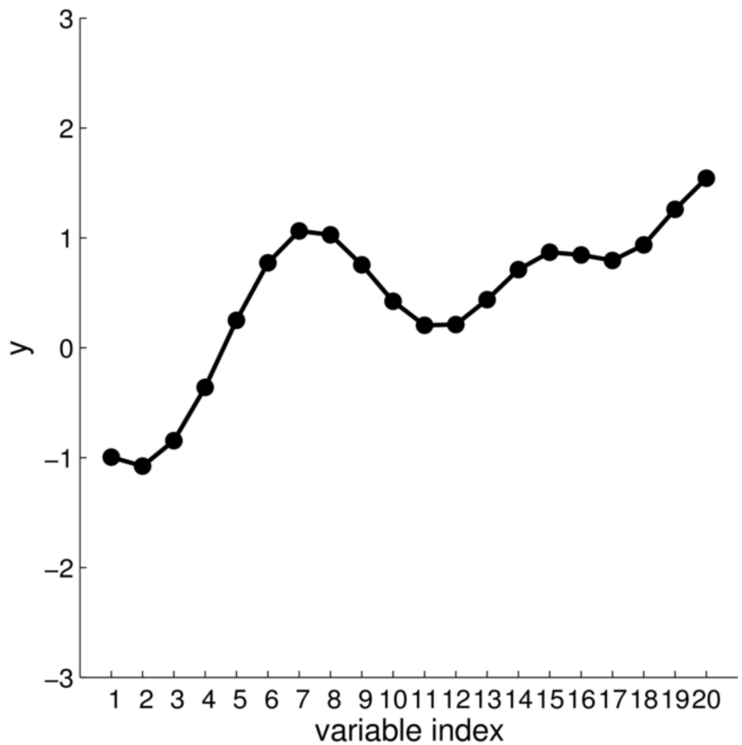


Visualizing Gaussian Processes

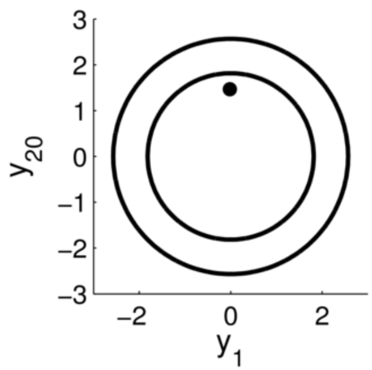


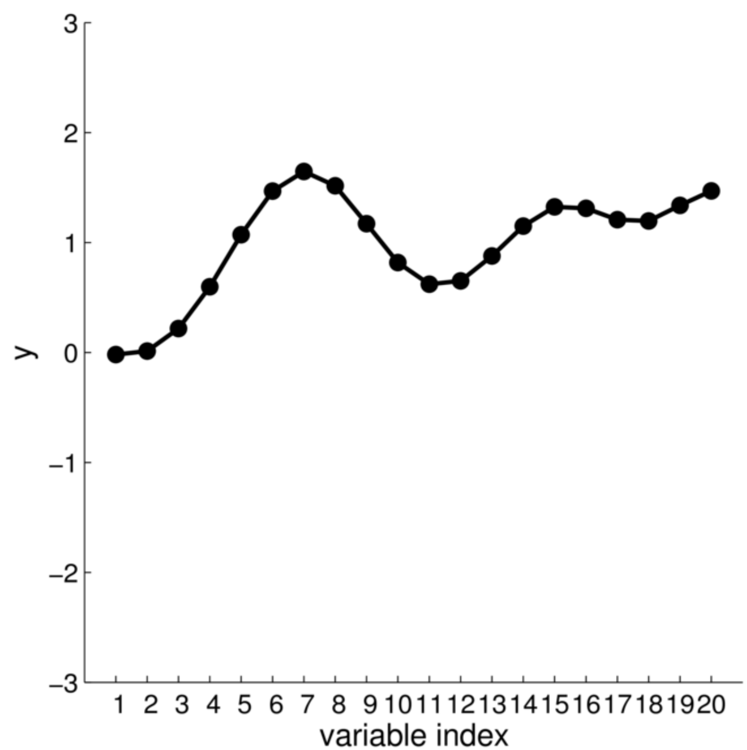
$\Sigma =$



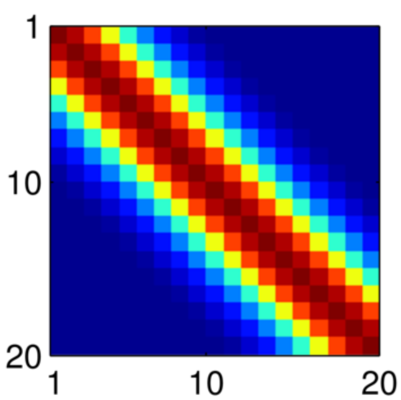


Visualizing Gaussian Processes

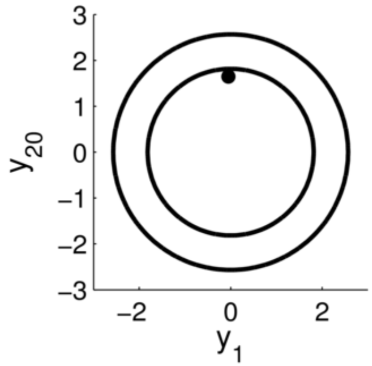


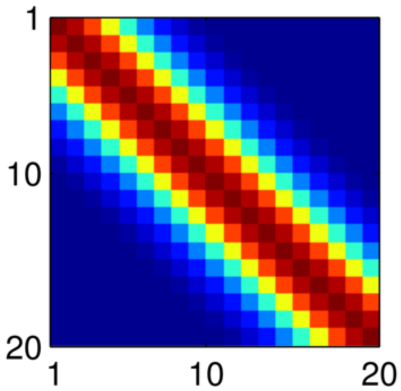


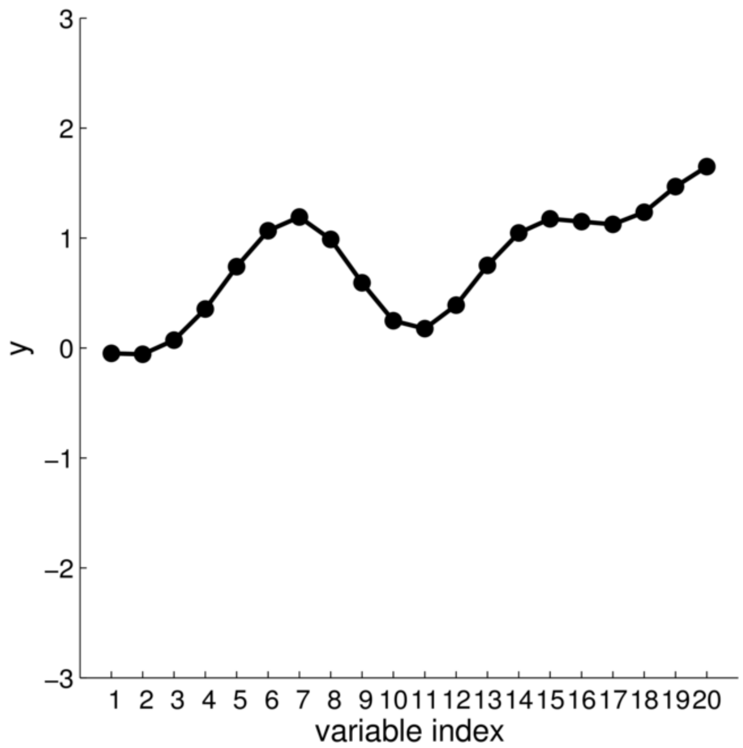
$\Sigma =$



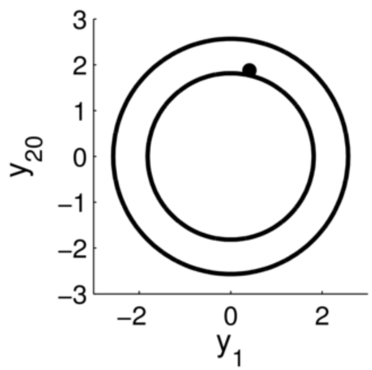
Visualizing Gaussian Processes

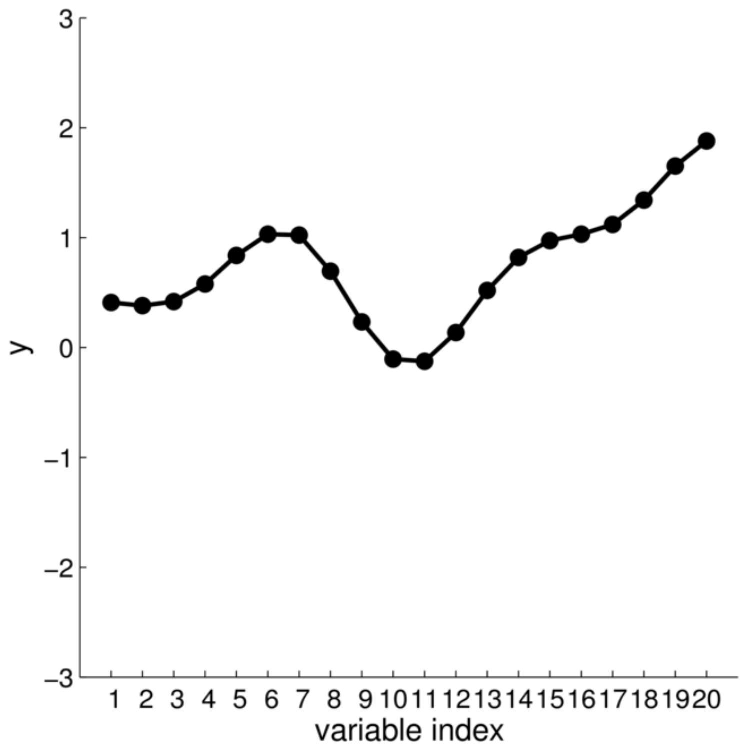


$\Sigma =$ 

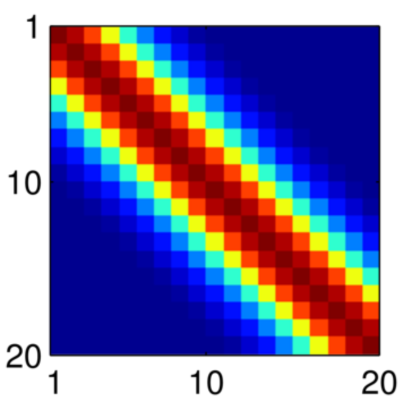


Visualizing Gaussian Processes

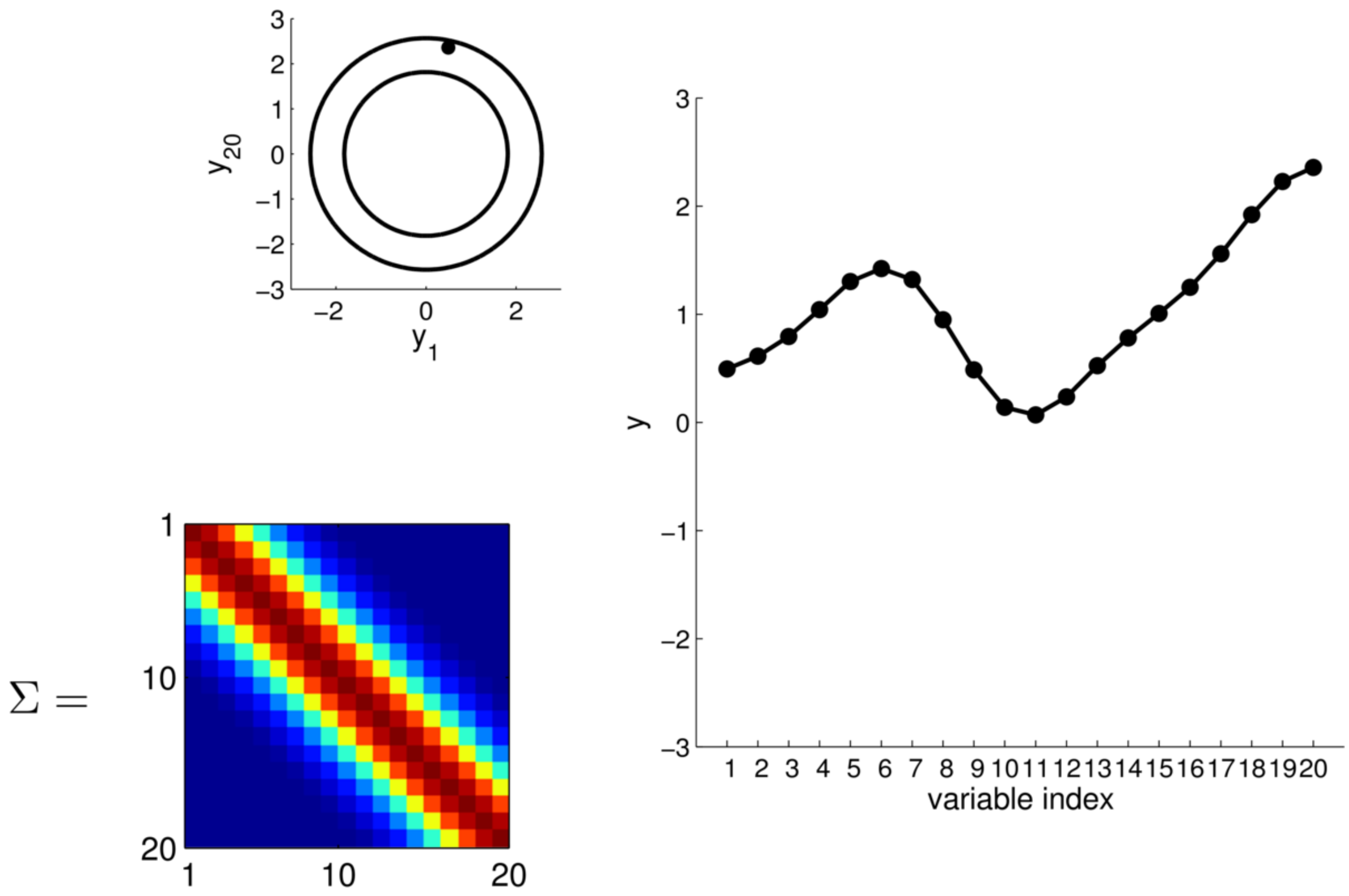




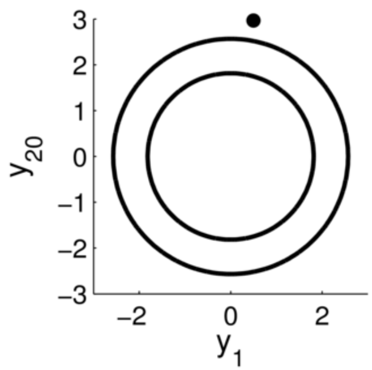
$\Sigma =$



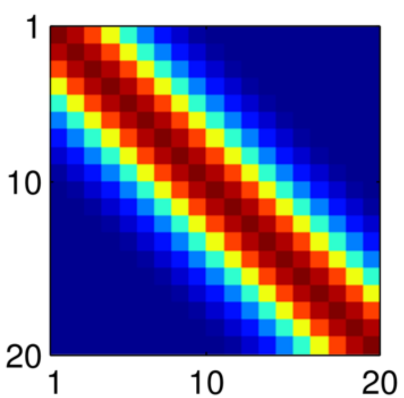
Visualizing Gaussian Processes

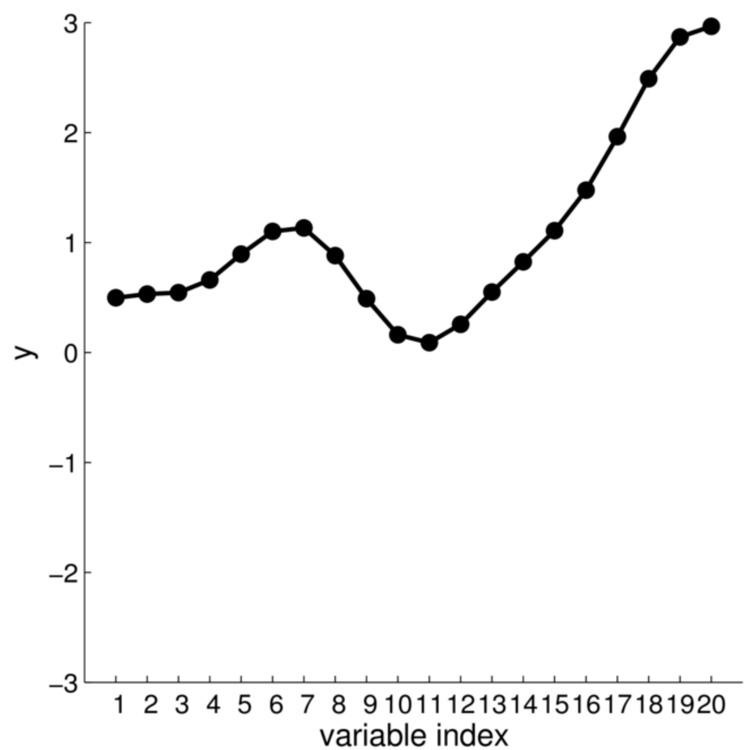


Visualizing Gaussian Processes

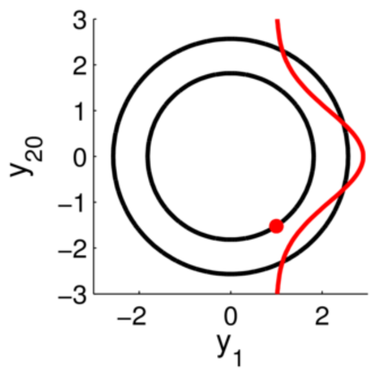


$\Sigma =$

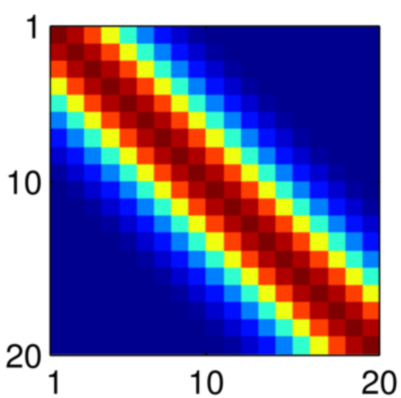


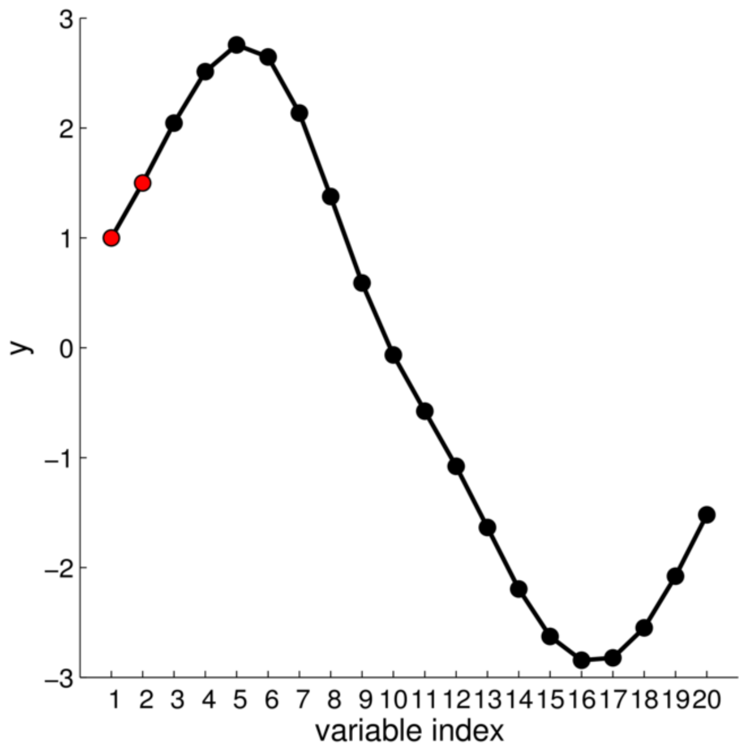


Visualizing Gaussian Processes

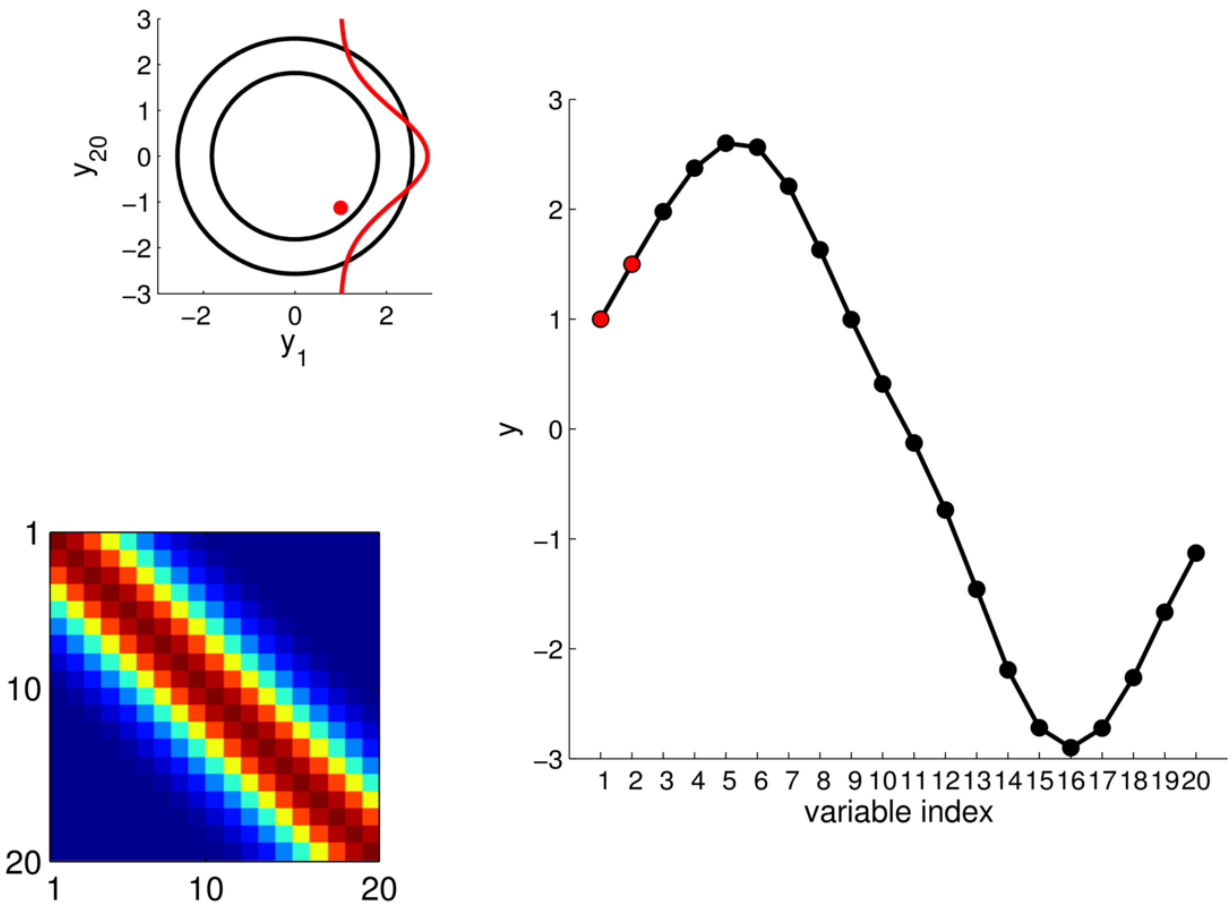


$\Sigma =$





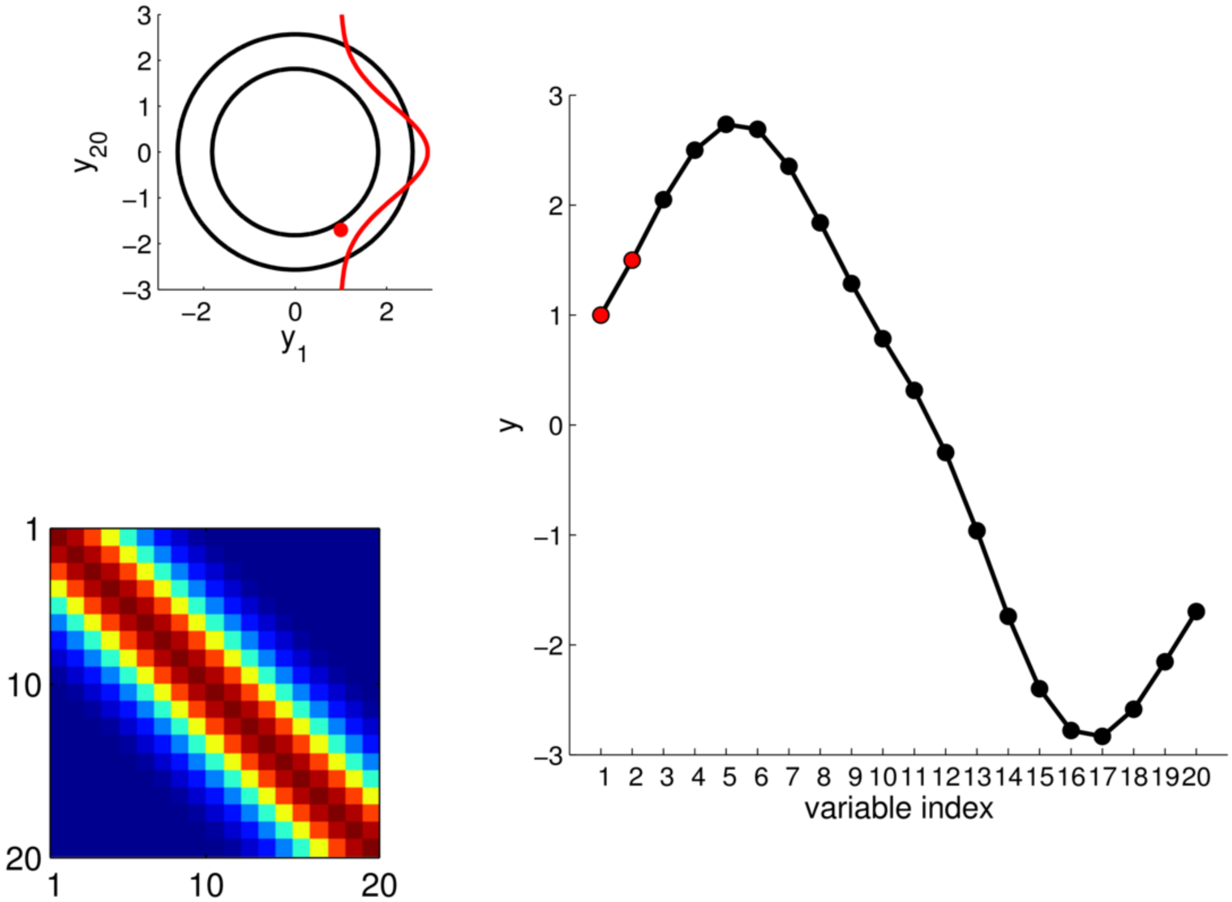
Visualizing Gaussian Processes



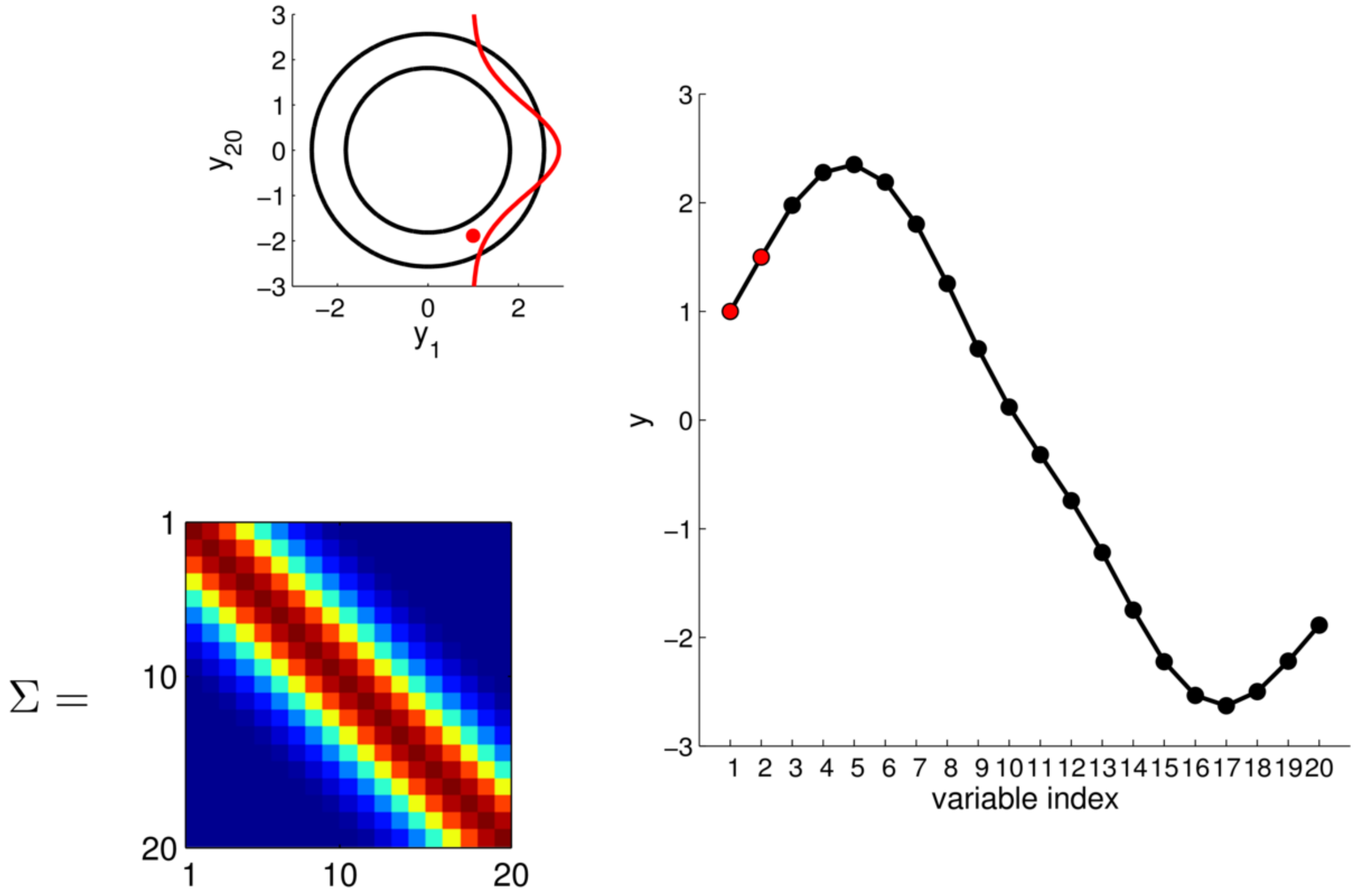
$\Sigma =$

Visualizing Gaussian Processes

$\Sigma =$

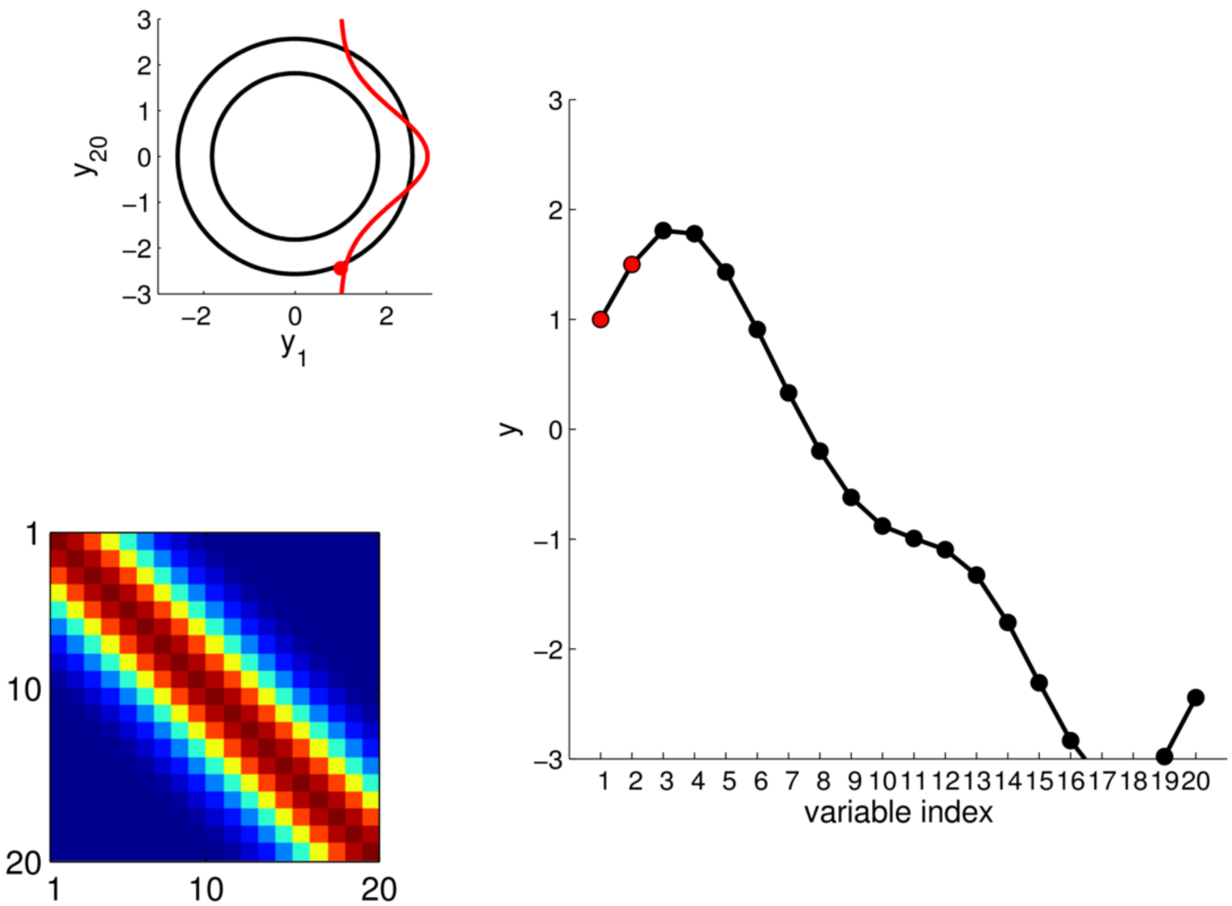


Visualizing Gaussian Processes



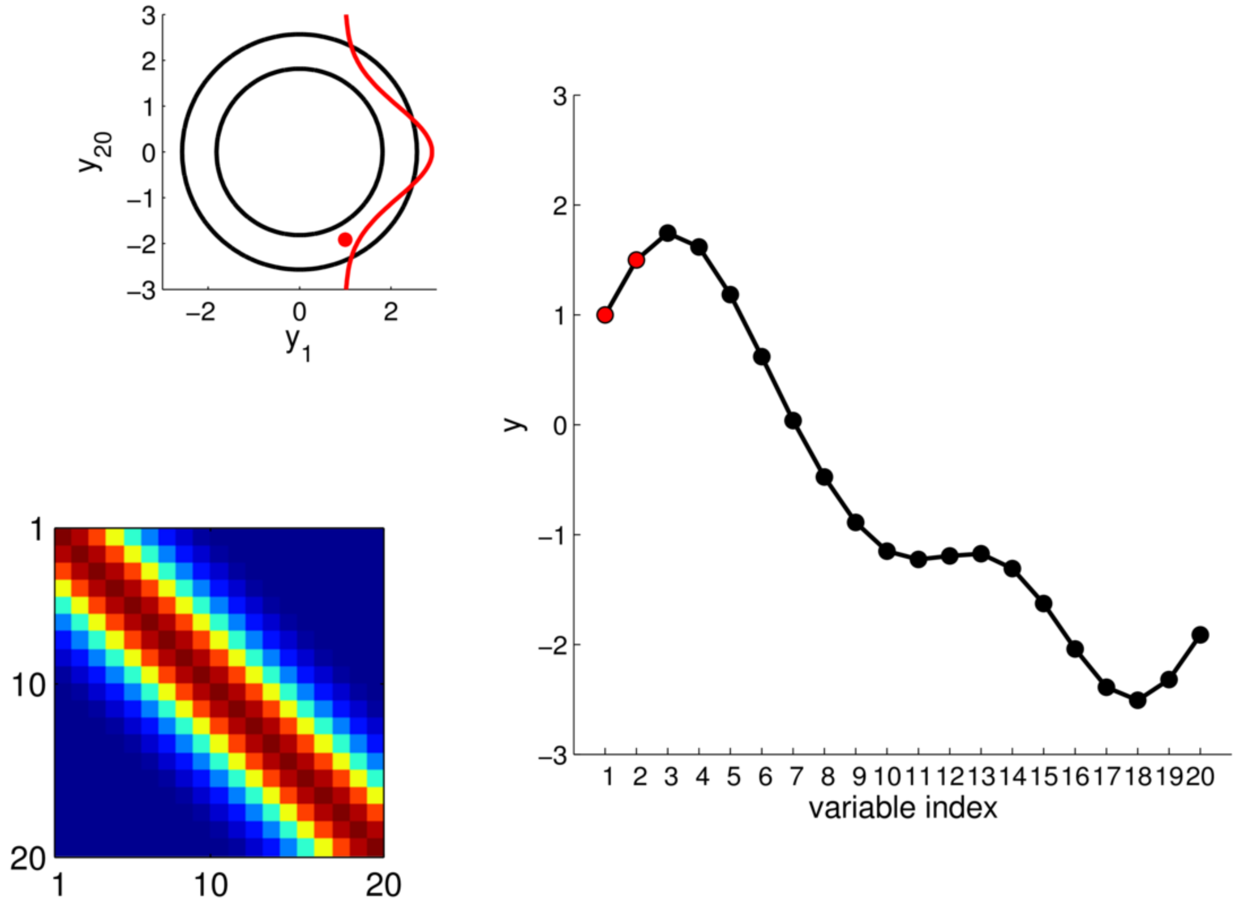
Visualizing Gaussian Processes

$\Sigma =$

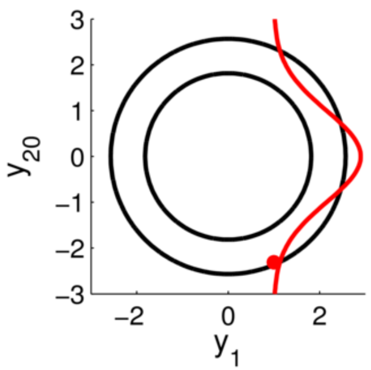


Visualizing Gaussian Processes

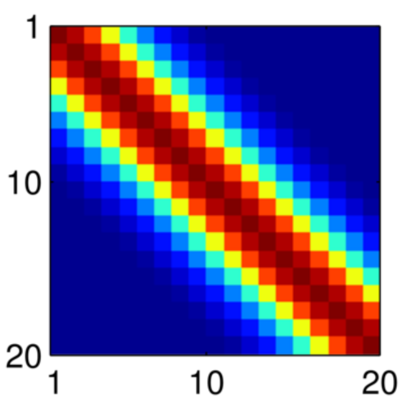
$\Sigma =$

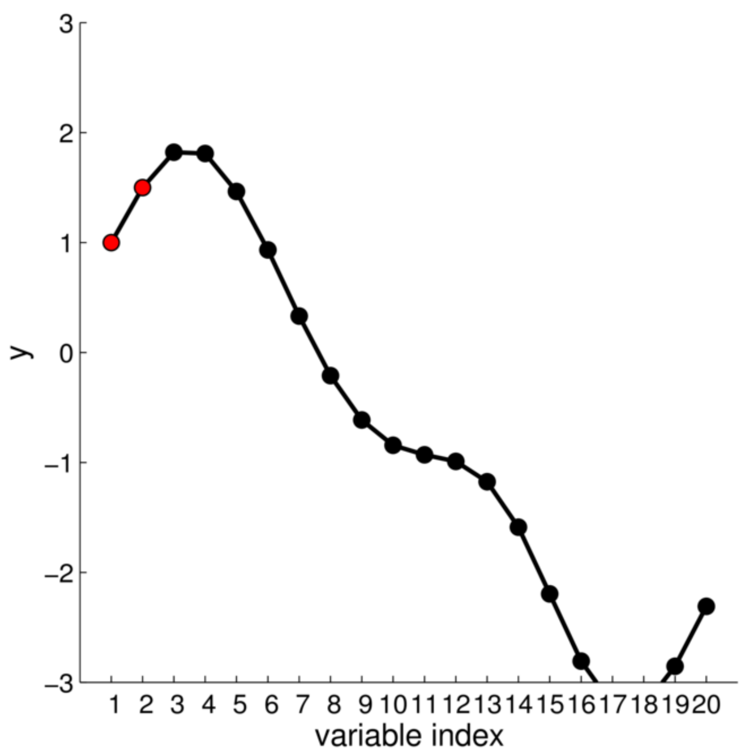


Visualizing Gaussian Processes

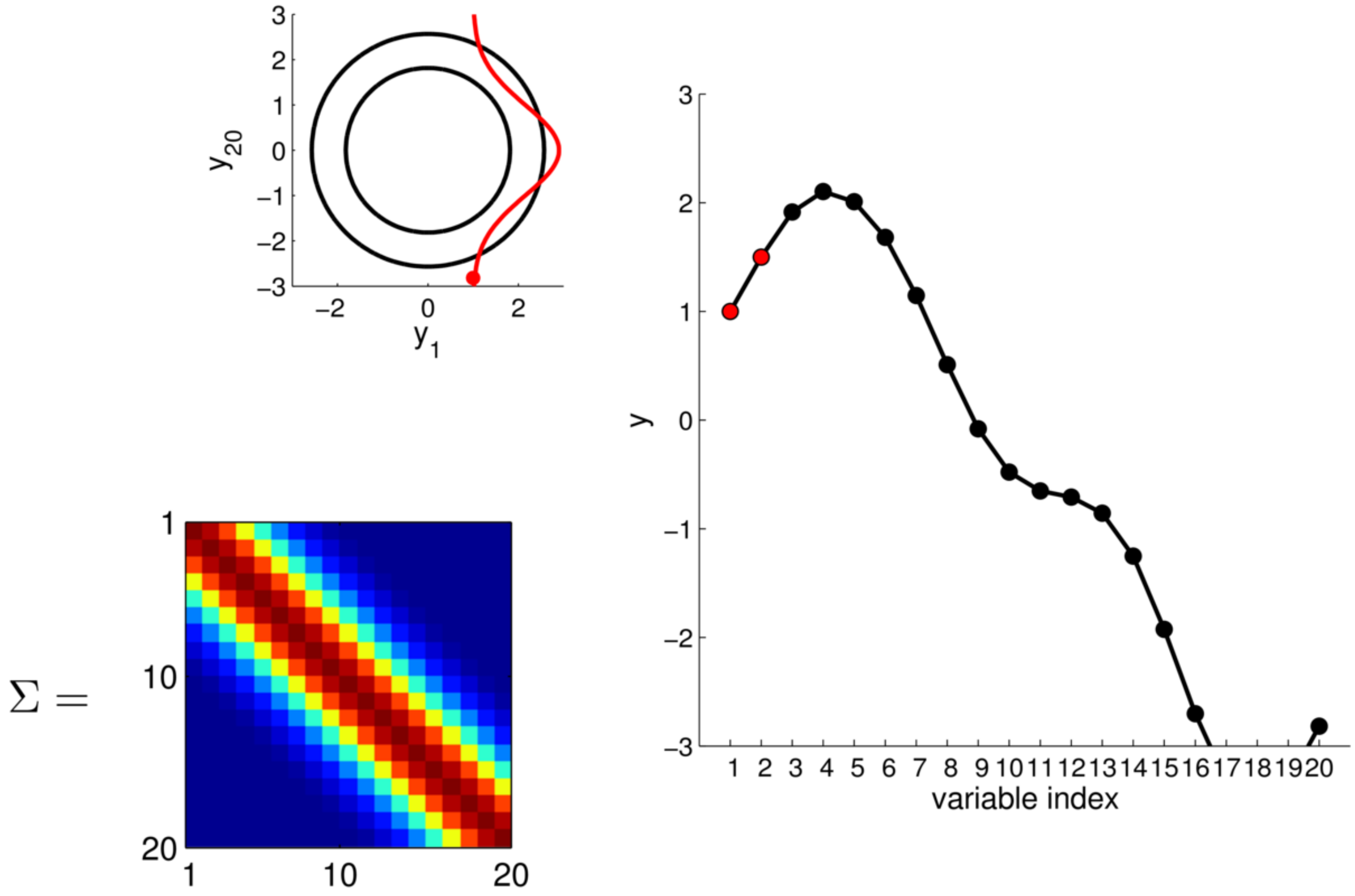


$\Sigma =$

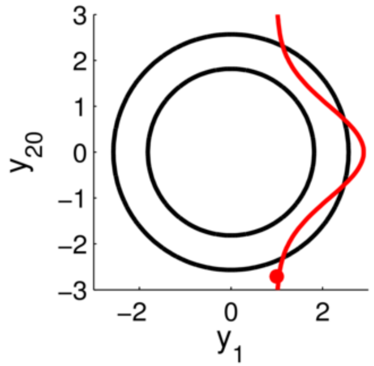


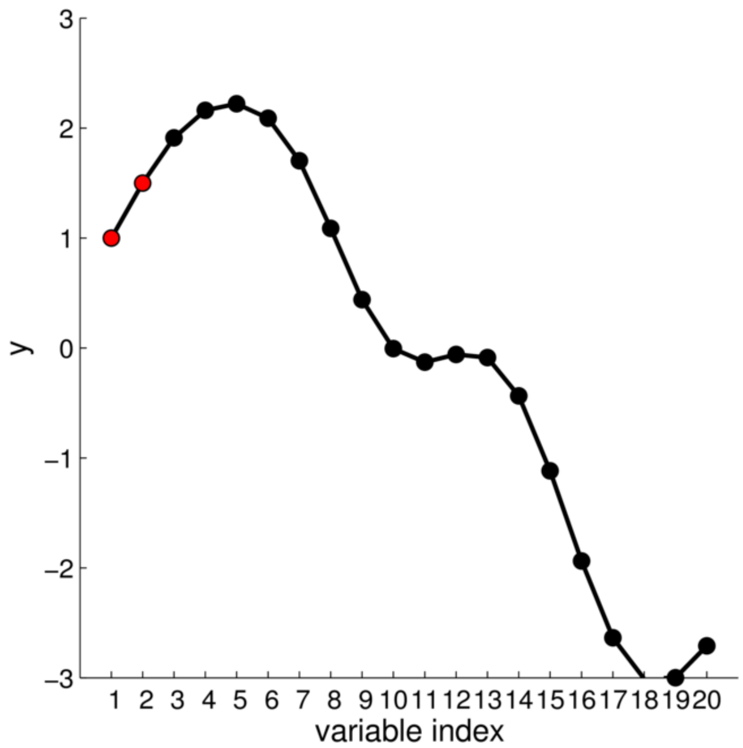


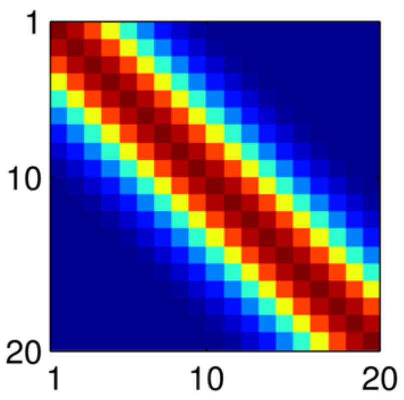
Visualizing Gaussian Processes



Visualizing Gaussian Processes

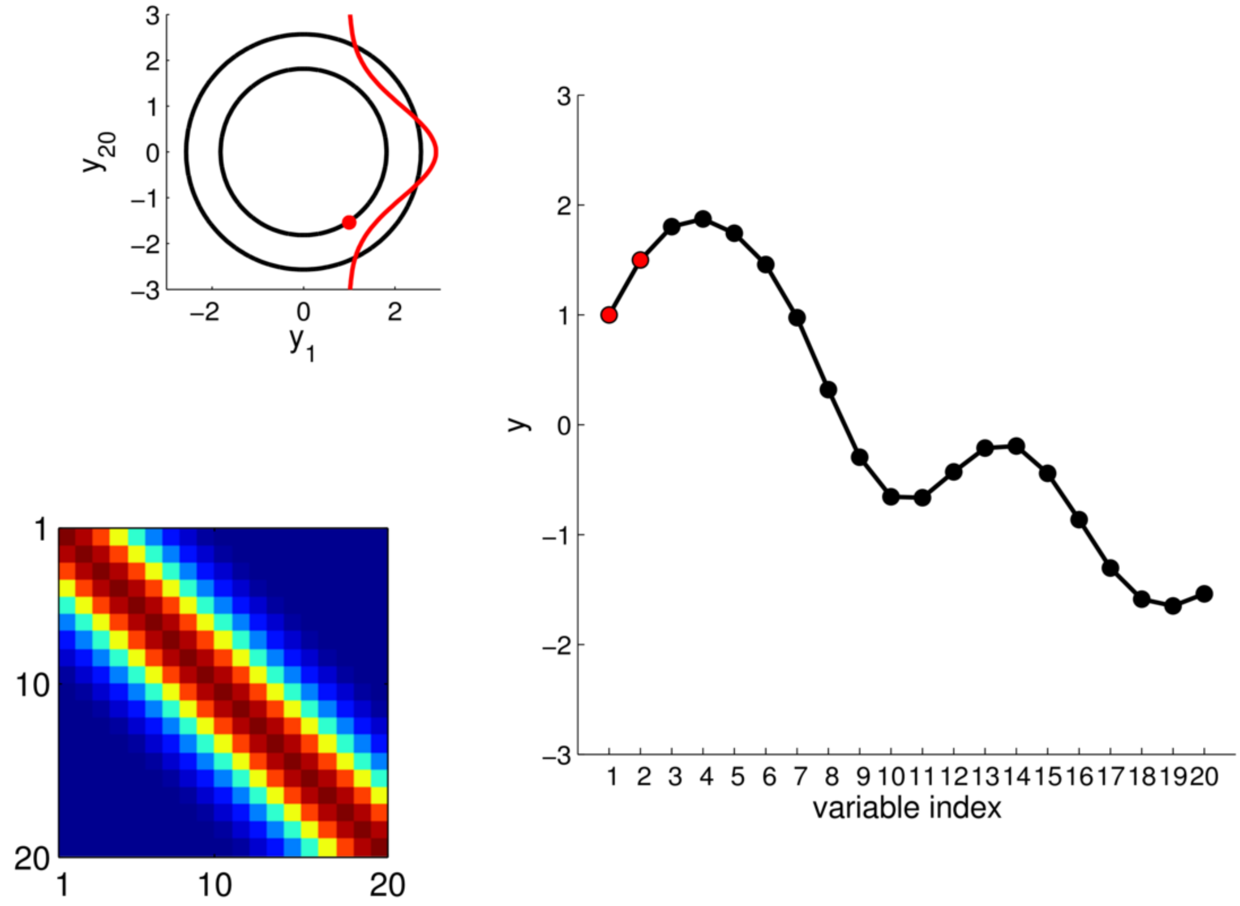




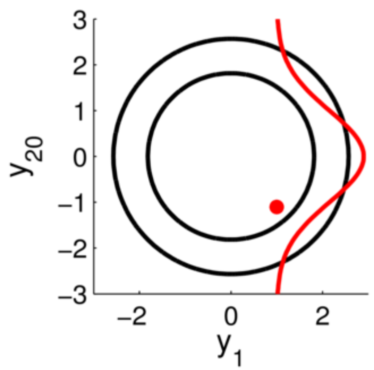
$\Sigma =$ 

Visualizing Gaussian Processes

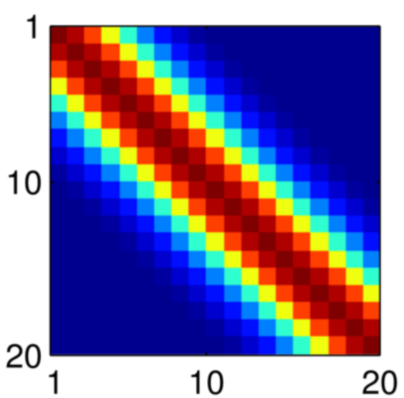
$\Sigma =$

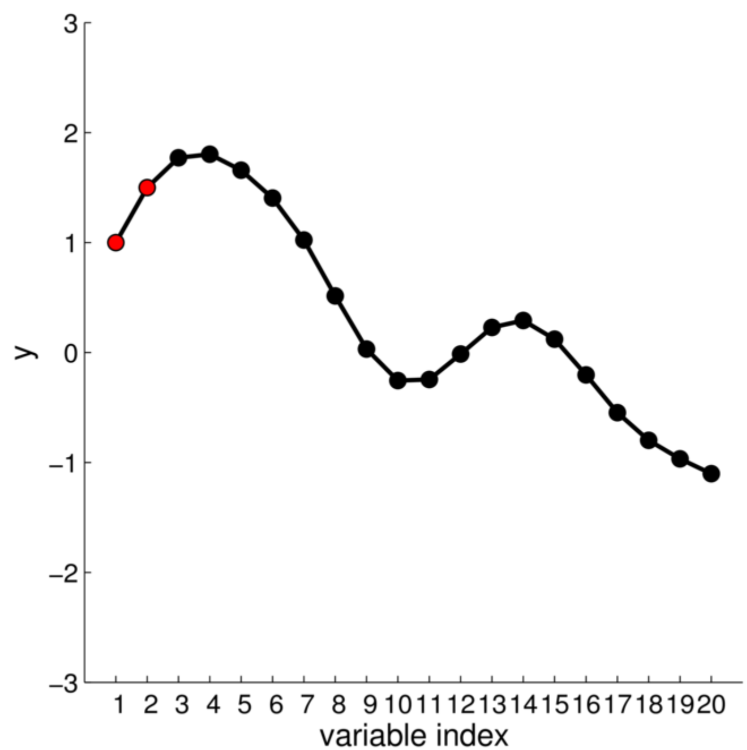


Visualizing Gaussian Processes



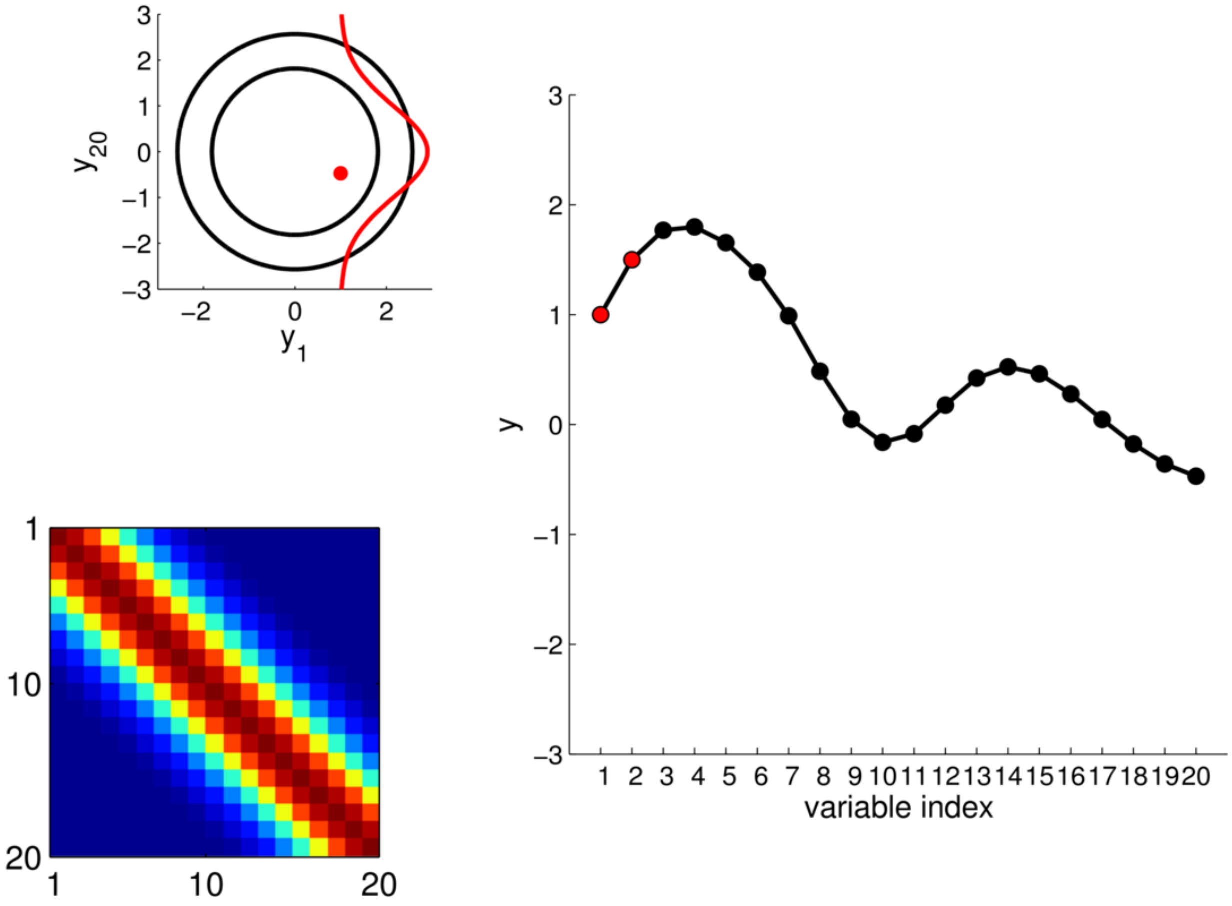
$\Sigma =$



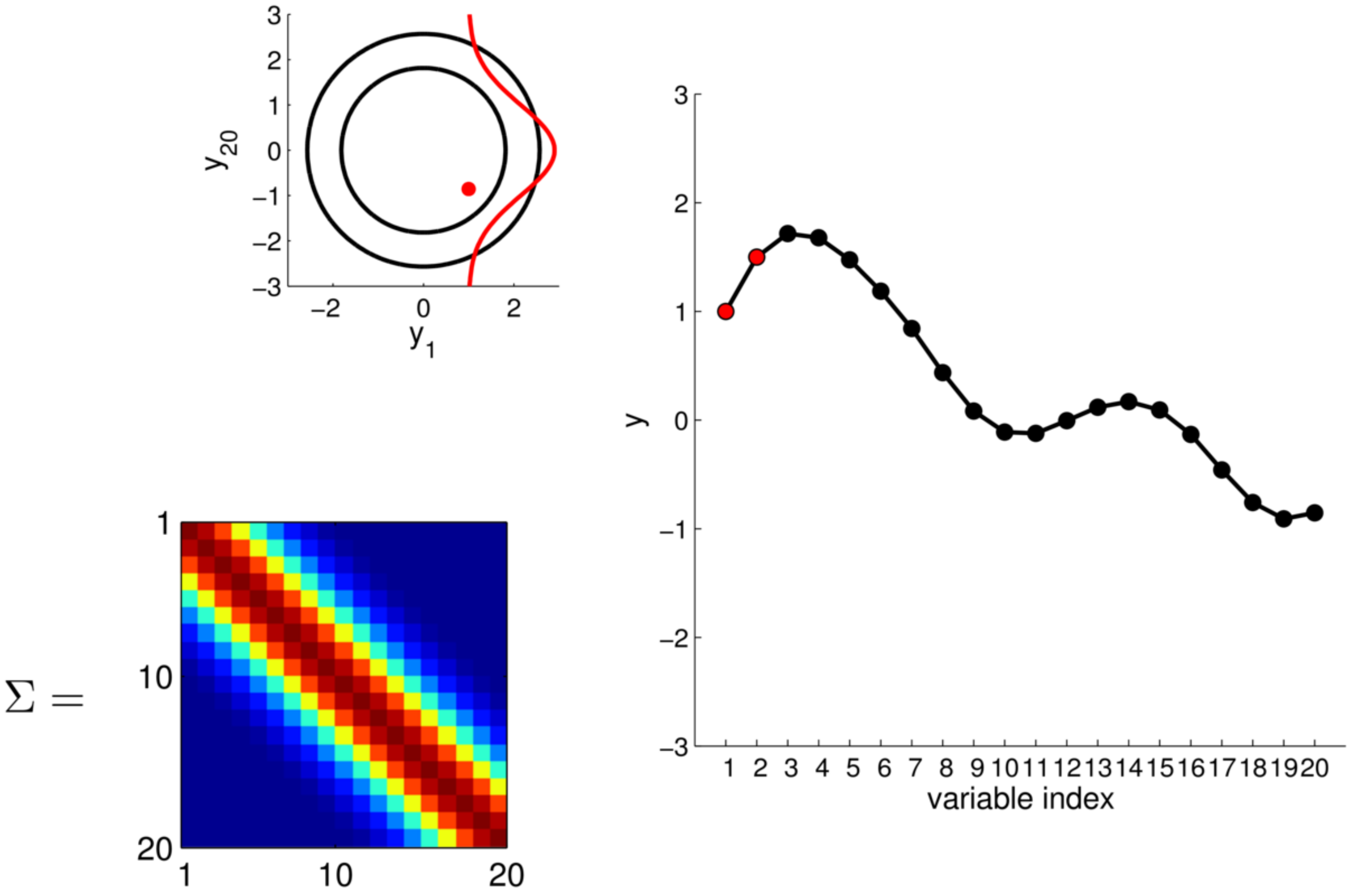


Visualizing Gaussian Processes

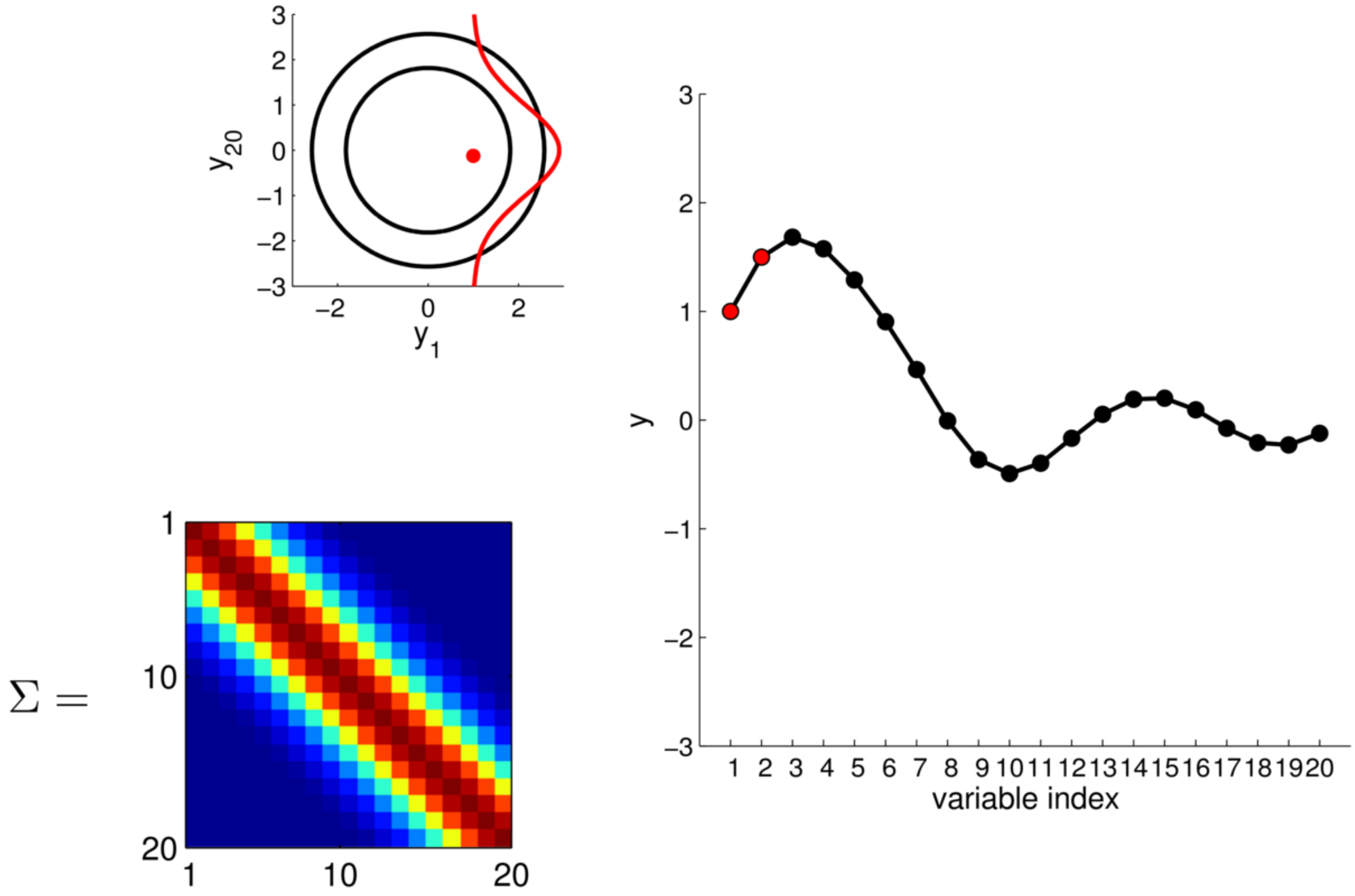
$\Sigma =$



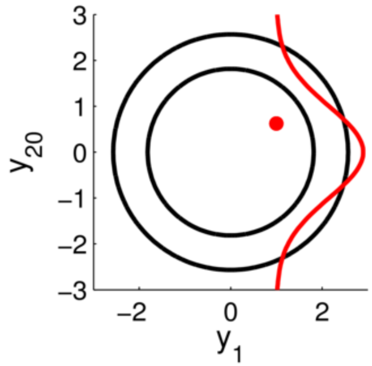
Visualizing Gaussian Processes



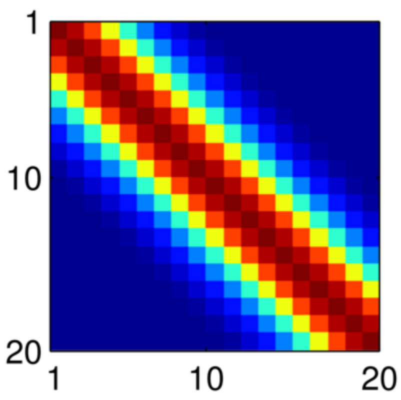
Visualizing Gaussian Processes

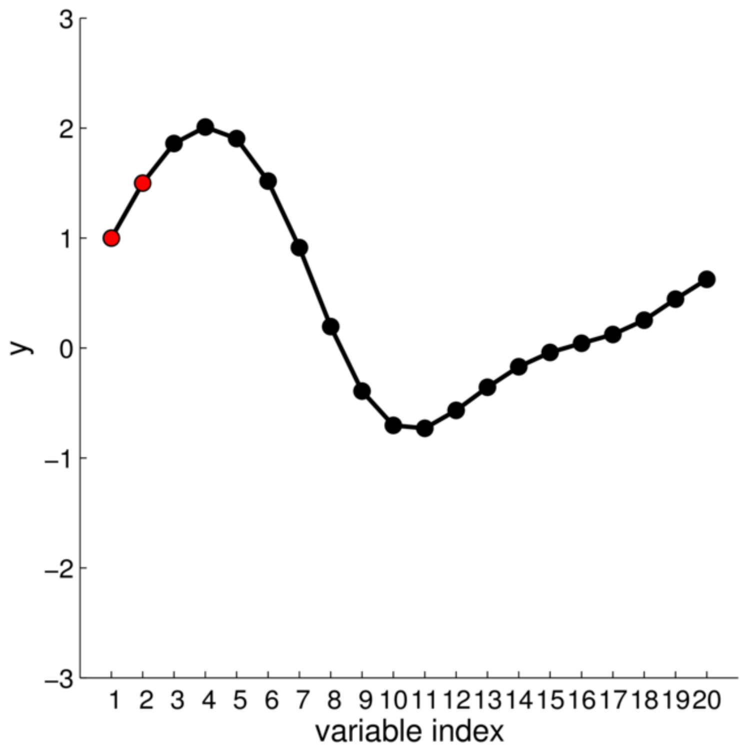


Visualizing Gaussian Processes



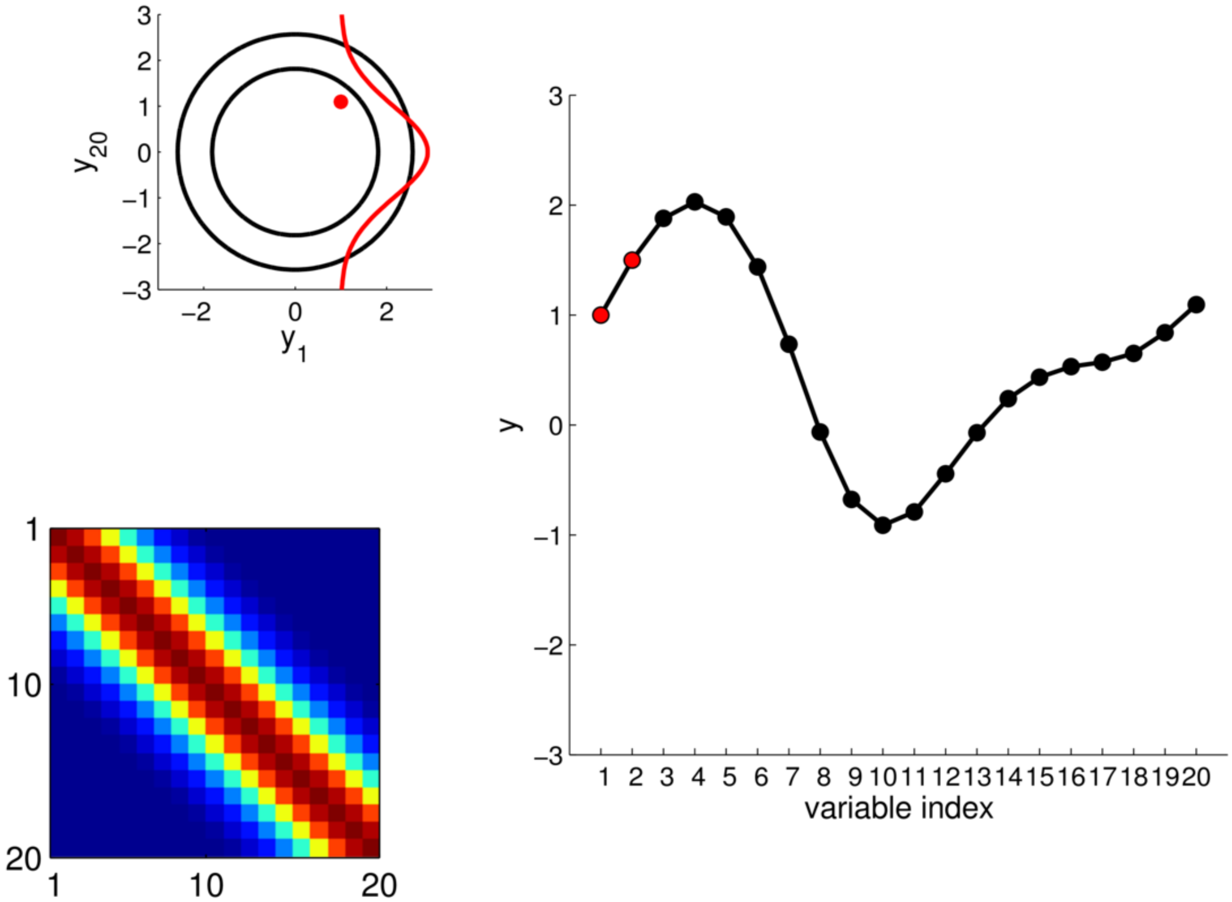
$\Sigma =$



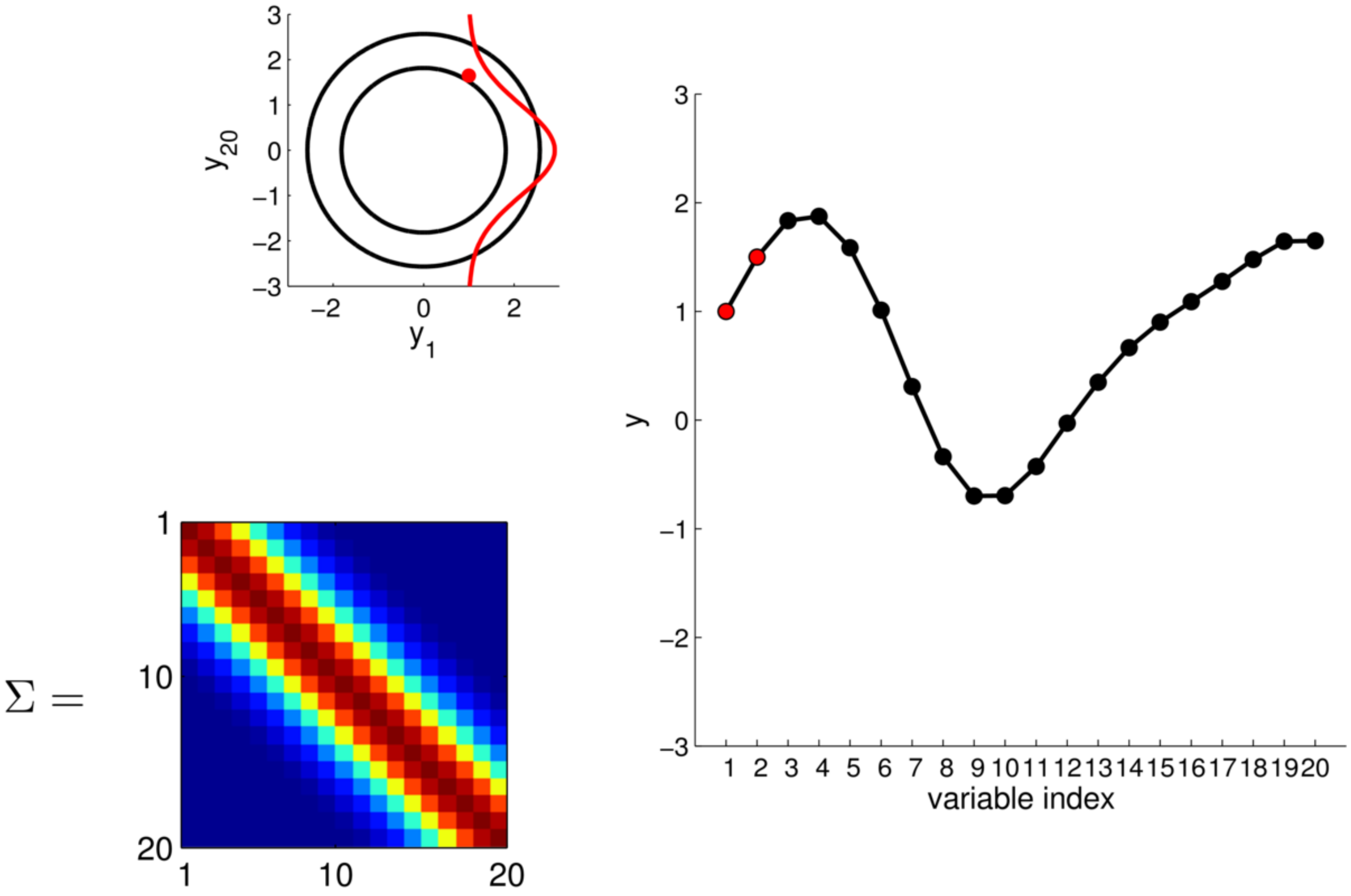


Visualizing Gaussian Processes

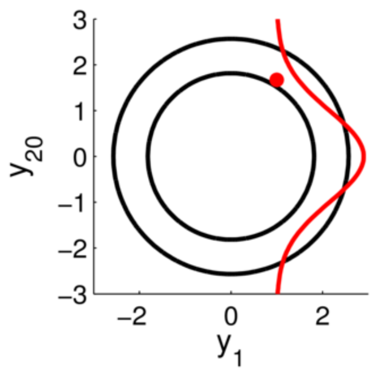
$\Sigma =$

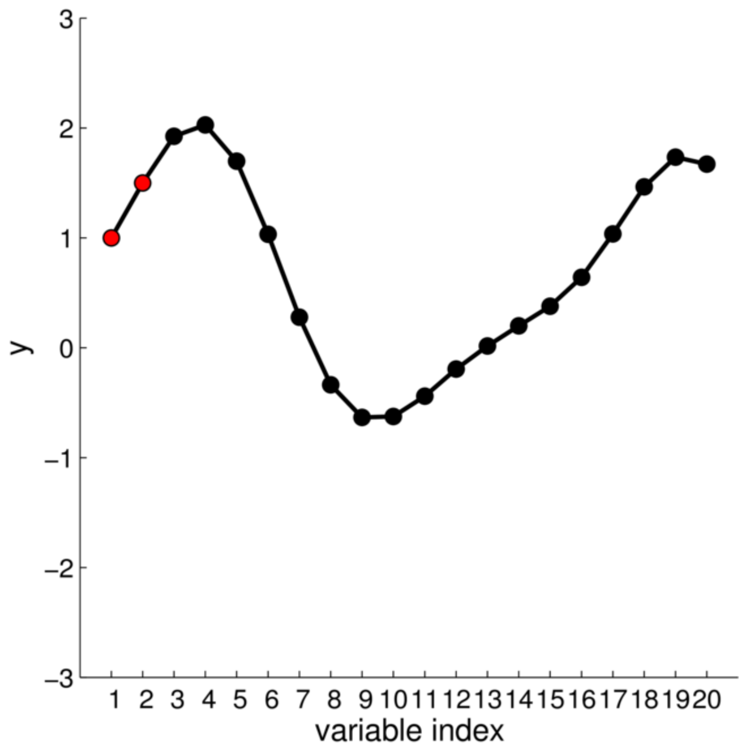


Visualizing Gaussian Processes

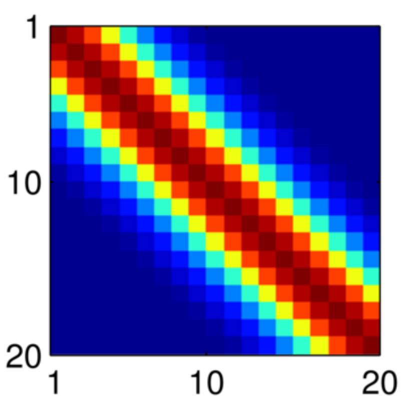


Visualizing Gaussian Processes

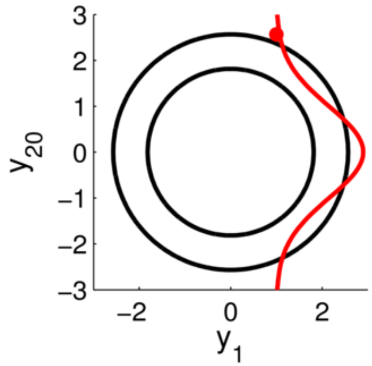


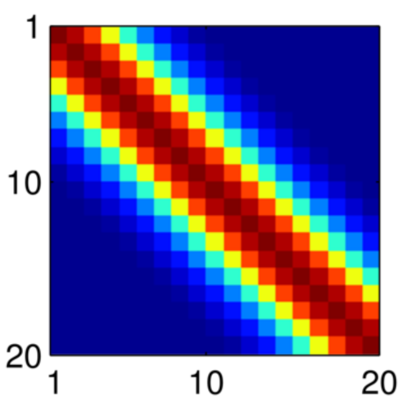


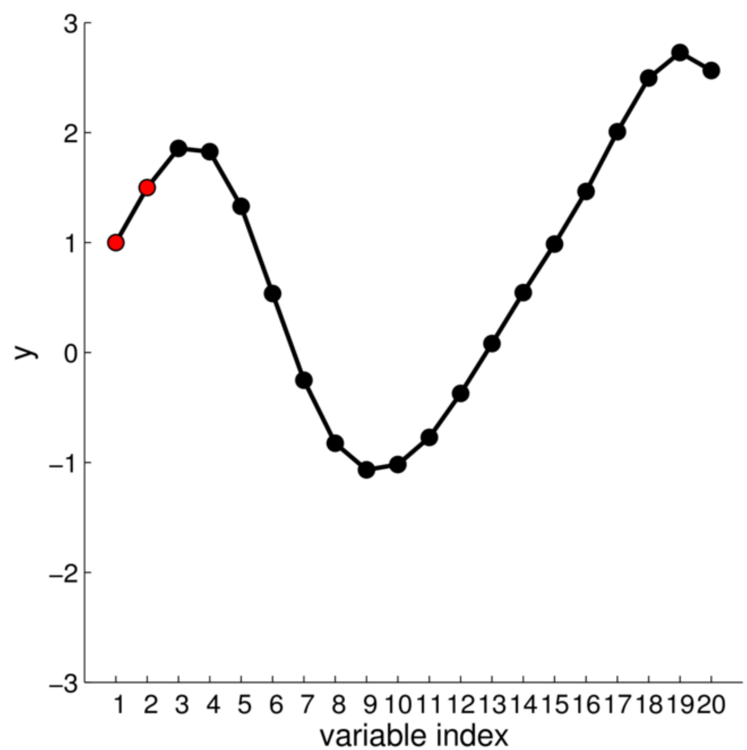
$\Sigma =$



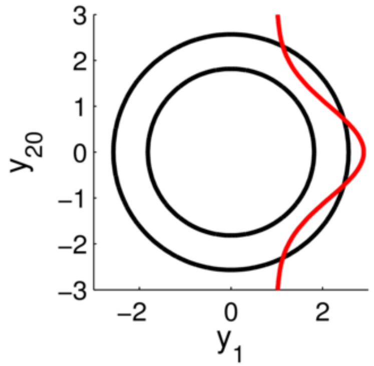
Visualizing Gaussian Processes



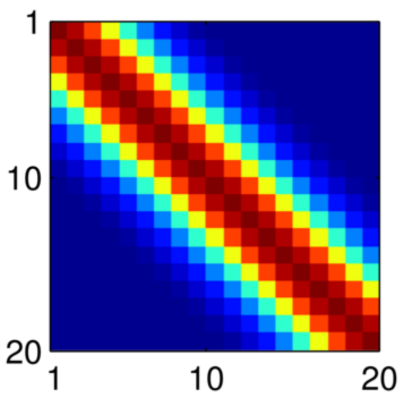
$\Sigma =$ 

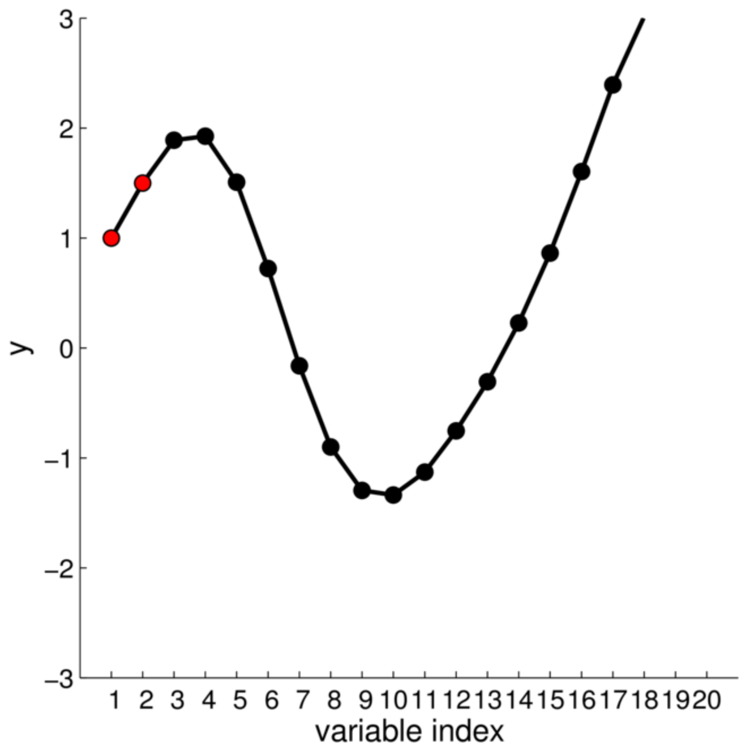


Visualizing Gaussian Processes



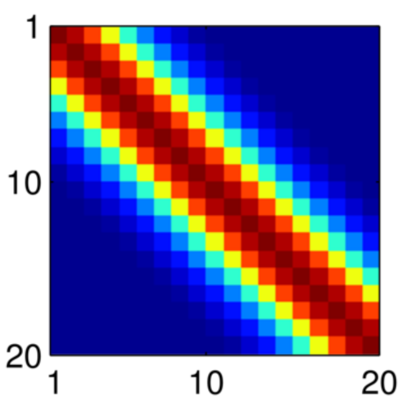
$\Sigma =$

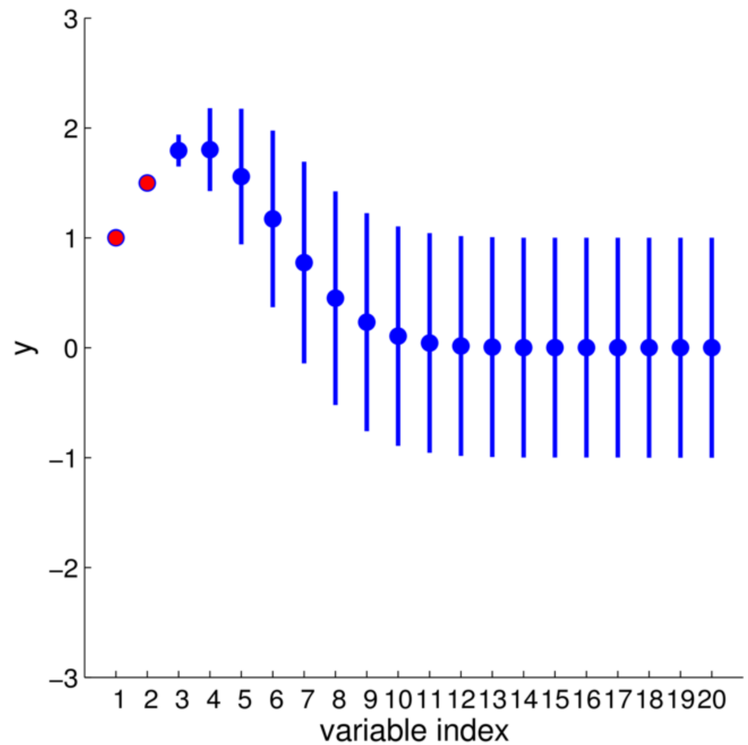




Regression via Gaussian Process

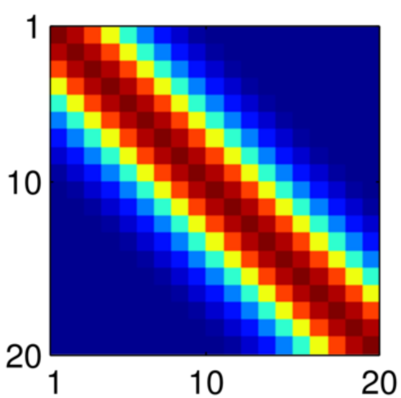
$\Sigma =$

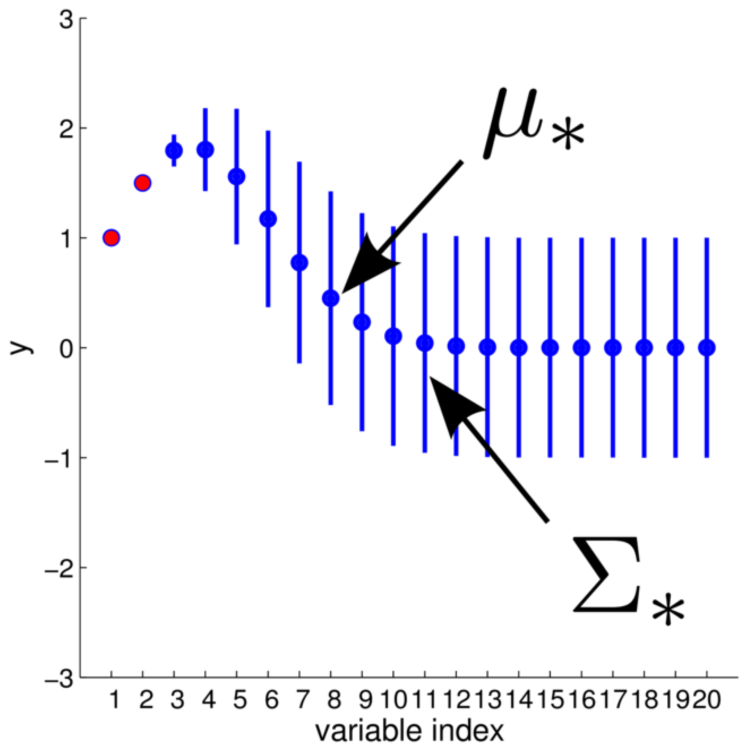




Regression via Gaussian Process

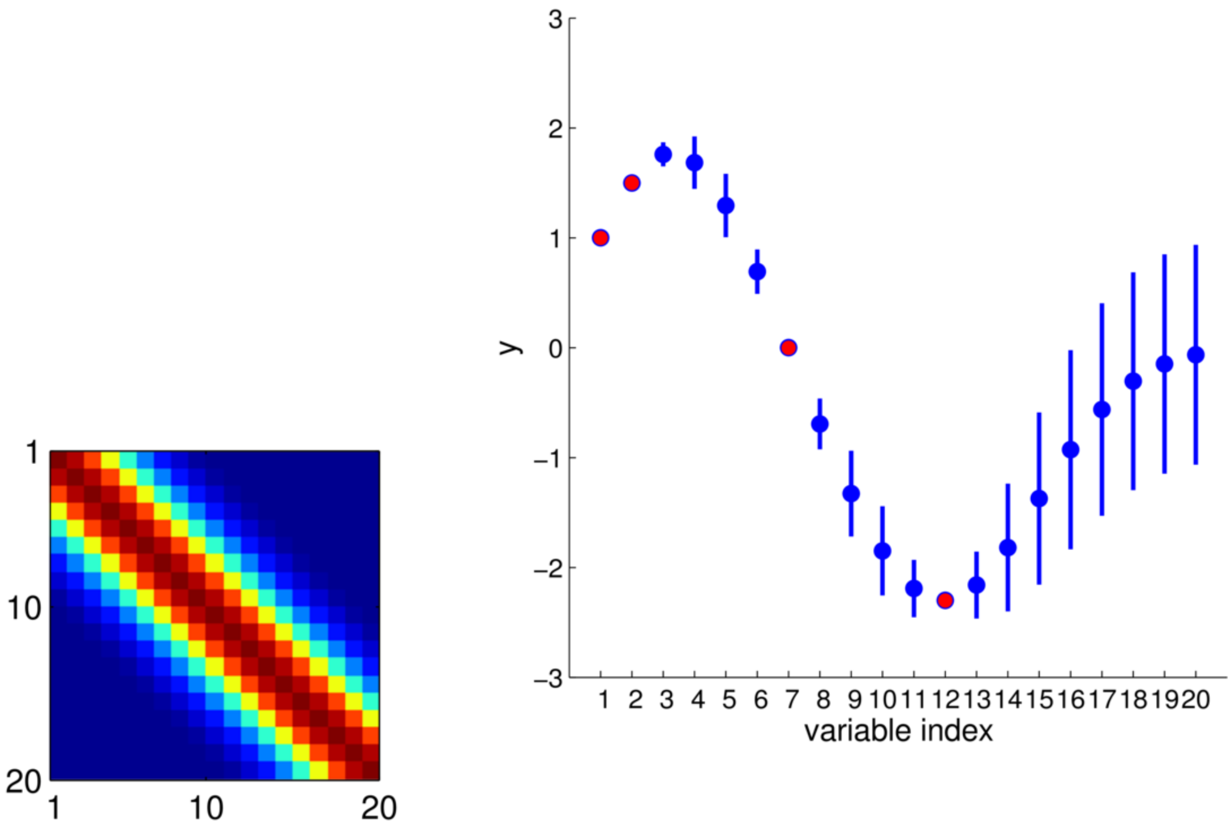
$\Sigma =$





Regression via Gaussian Process

$\Sigma =$



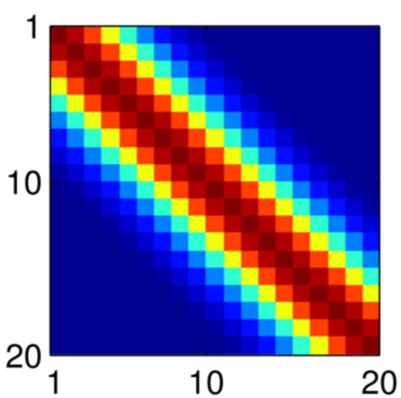
Regression via Gaussian Process

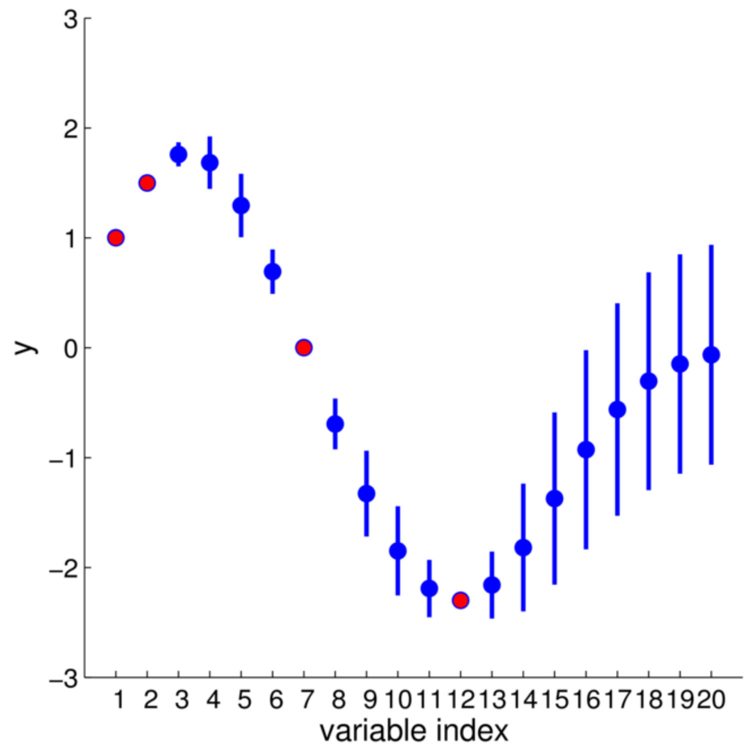
I: the Identity matrix

$$\Sigma(x_1, x_2) = K(x_1, x_2) + I\sigma_y^2$$

$$K(x_1, x_2) = \sigma^2 e^{-\frac{1}{2l^2}(x_1 - x_2)^2}$$

l: constant for horizontal scale

$\Sigma =$ 

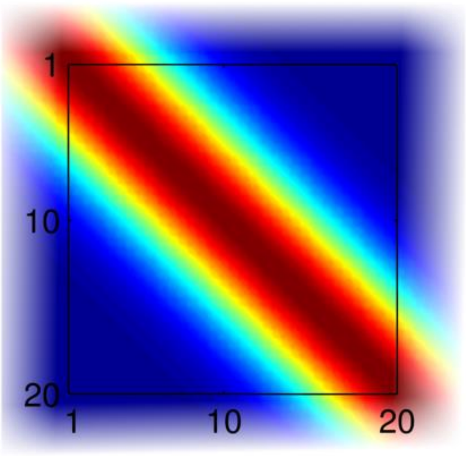


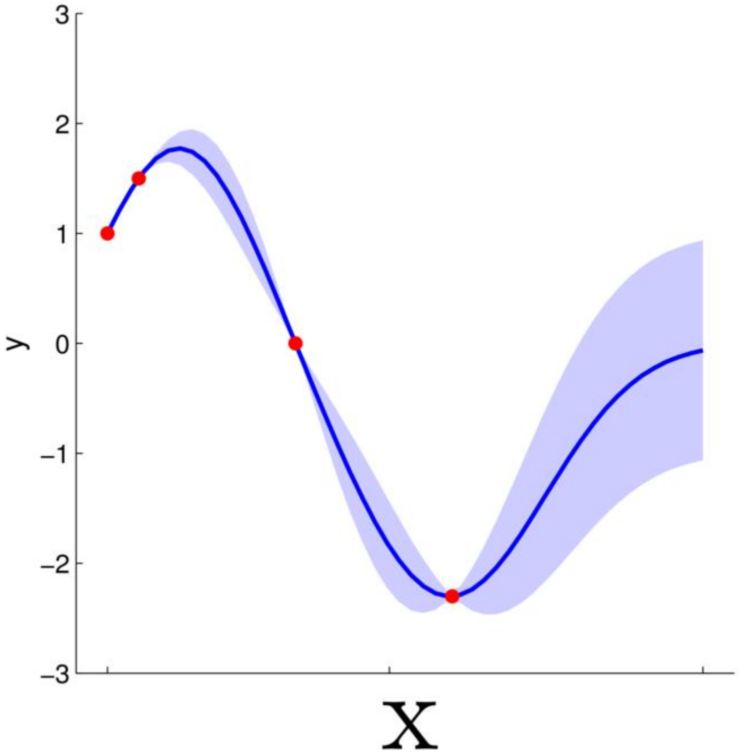
Probabilistic Inference in Function Space

$$\Sigma(x_1, x_2) = K(x_1, x_2) + I\sigma_y^2$$

$$K(x_1, x_2) = \sigma^2 e^{-\frac{1}{2l^2}(x_1 - x_2)^2}$$

$\Sigma =$





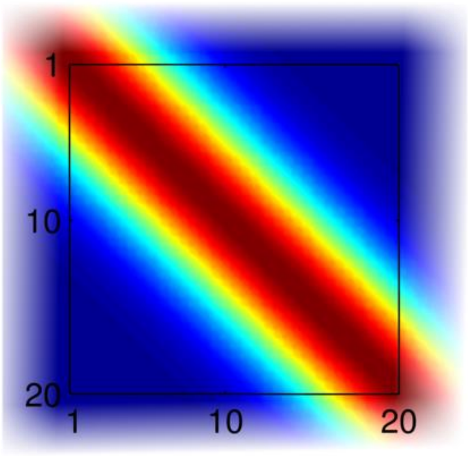
Probabilistic Inference in Function Space

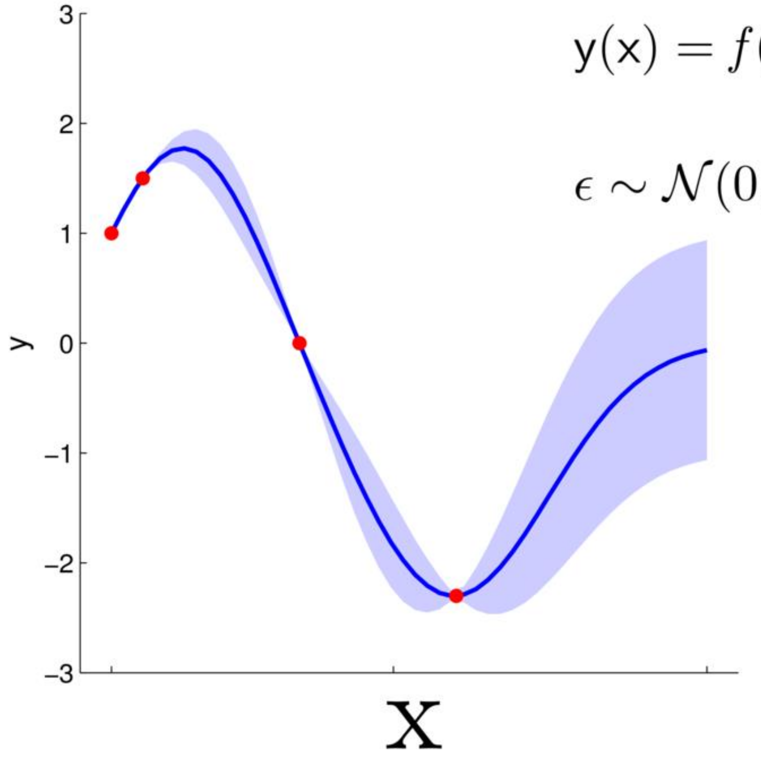
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

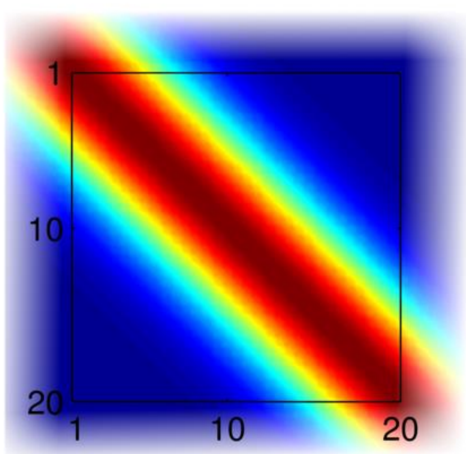
Probabilistic Inference in Function Space

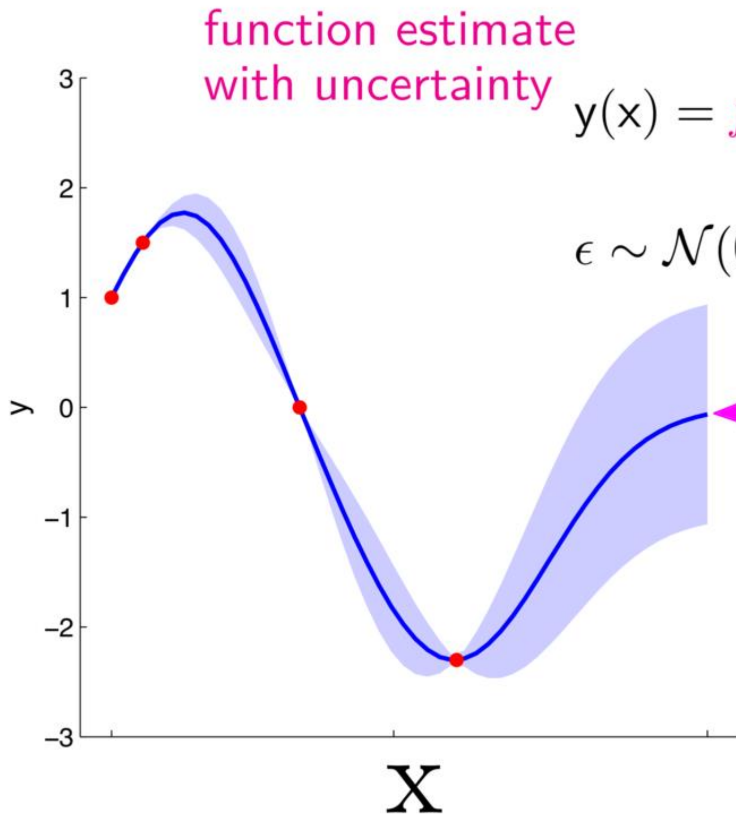
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Probabilistic Inference in Function Space

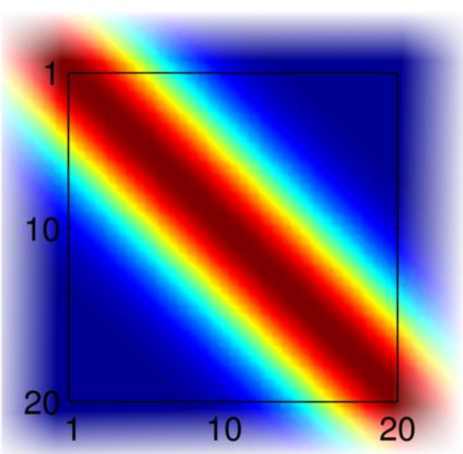
Non-parametric (∞-parametric)

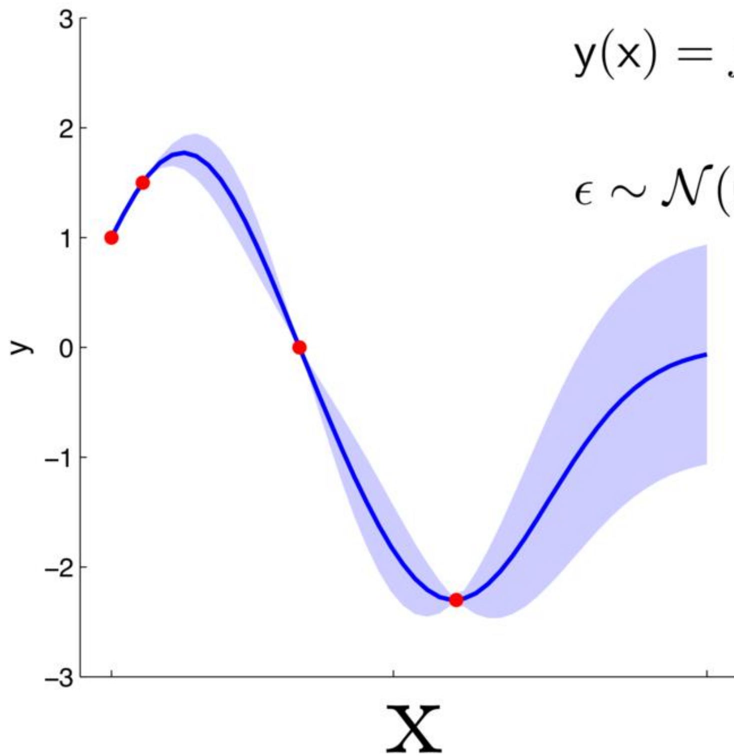
$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

noise

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Probabilistic Inference in Function Space

Non-parametric (∞-parametric)

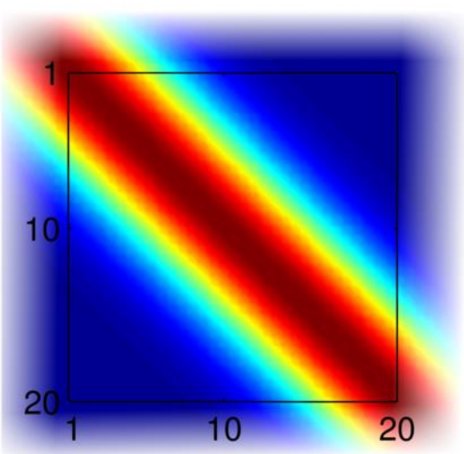
$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

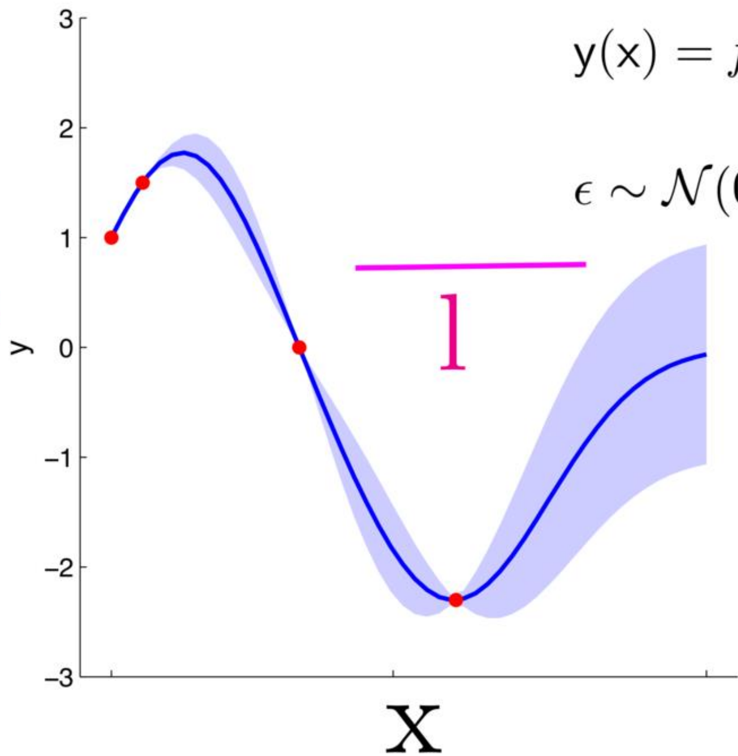
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

horizontal-scale

$\Sigma =$





Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Probabilistic Inference in Function Space

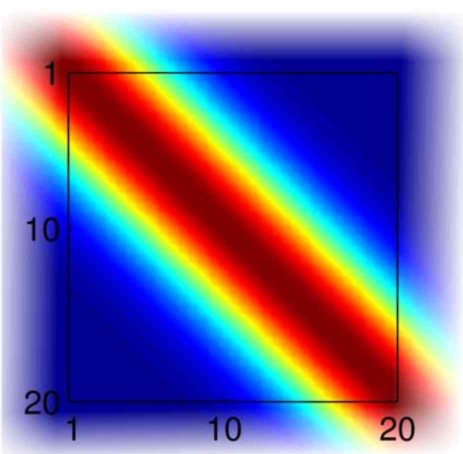
Non-parametric (∞-parametric)

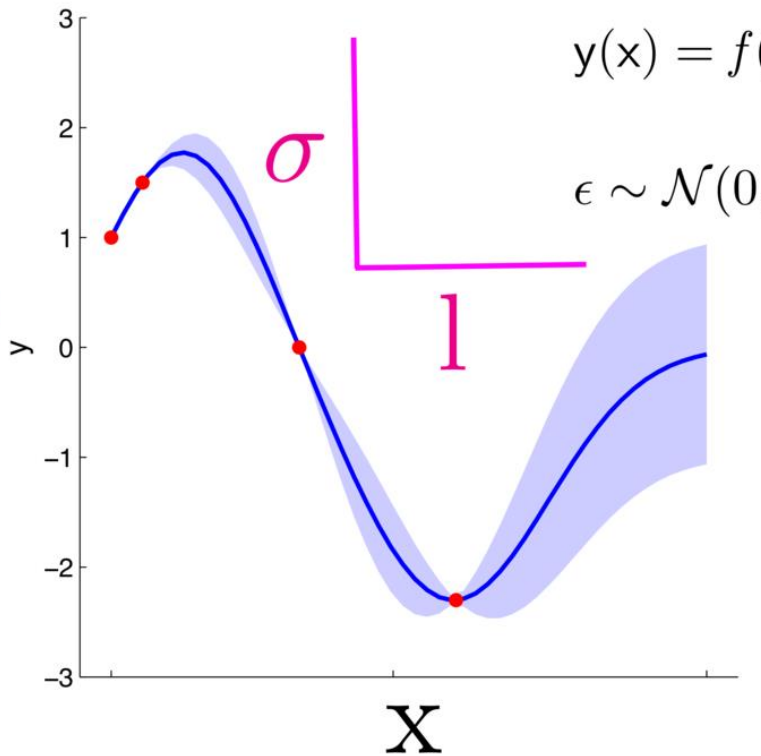
$$p(y|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(x_1, x_2) = K(x_1, x_2) + I\sigma_y^2$$

$$K(x_1, x_2) = \sigma^2 e^{-\frac{1}{2l^2}(x_1 - x_2)^2}$$

vertical-scale horizontal-scale

$\Sigma =$ 



Parametric model

$$y(x) = f(x; \theta) + \sigma_y \epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Mathematical Foundations: Definition

Gaussian process = generalisation of multivariate Gaussian distribution to infinitely many variables.

Definition: a Gaussian process is a collection of random variables, any finite number of which have (consistent) Gaussian distributions.

A Gaussian distribution is fully specified by a mean vector, $\boldsymbol{\mu}$, and covariance matrix Σ:

$$\mathbf{f} = (\mathsf{f}_1, \ldots, \mathsf{f}_n) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma), \quad \text{indices} \quad i = 1, \ldots, n$$

A Gaussian process is fully specified by a mean function $m(\mathbf{x})$ and covariance function $\mathsf{K}(\mathbf{x}, \mathbf{x}')$:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), \mathsf{K}(\mathbf{x}, \mathbf{x}')\right), \quad \text{indices} \quad \mathbf{x}$$

Mathematical Foundations: Regression

Q1. What's the formal justification for how we were using GPs for regression?

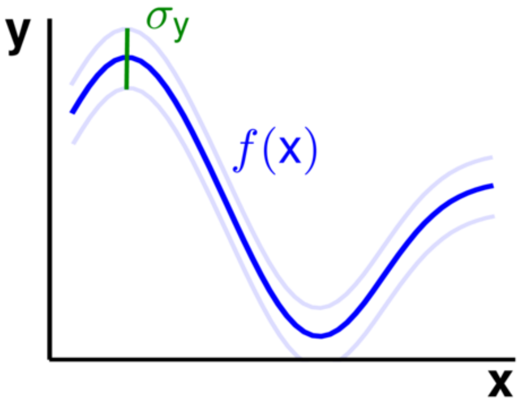
Mathematical Foundations: Regression

Q1. What's the formal justification for how we were using GPs for regression?

generative model (like non-linear regression)

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}) + \epsilon\sigma_{\mathrm{y}}$$

$$p(\epsilon) = \mathcal{N}(0, 1)$$



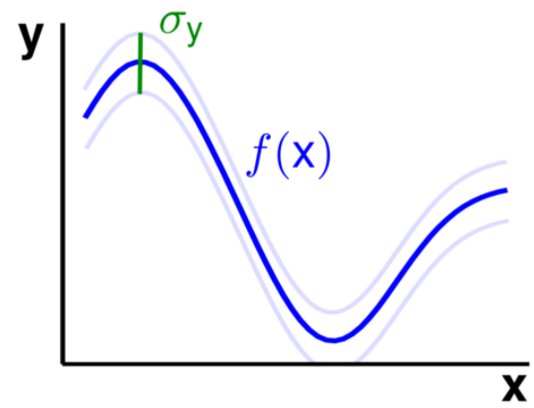
Mathematical Foundations: Regression

Q1. What's the formal justification for how we were using GPs for regression?

generative model (like non-linear regression)

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}) + \epsilon\sigma_{\mathsf{y}}$$

$$p(\epsilon) = \mathcal{N}(0, 1)$$



place GP prior over the non-linear function

$$p(f(\mathsf{x})|\theta) = \mathcal{GP}(0, \mathsf{K}(\mathsf{x}, \mathsf{x}'))$$

$$\mathsf{K}(\mathsf{x}, \mathsf{x}') = \sigma^2 \exp\left(-\frac{1}{2l^2}(\mathsf{x} - \mathsf{x}')^2\right)$$

(smoothly wiggling functions expected)

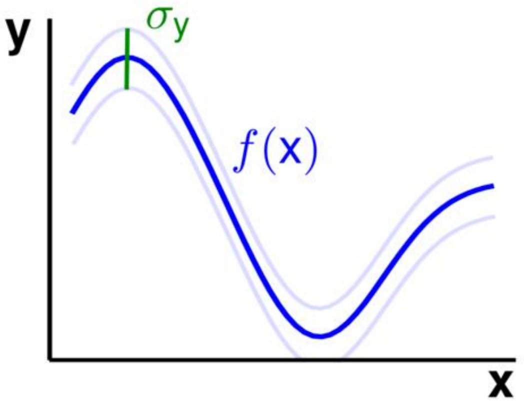
Mathematical Foundations: Regression

Q1. What's the formal justification for how we were using GPs for regression?

generative model (like non-linear regression)

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}) + \epsilon\sigma_{\mathsf{y}}$$

$$p(\epsilon) = \mathcal{N}(0, 1)$$



place GP prior over the non-linear function

$$p(f(\mathsf{x})|\theta) = \mathcal{GP}(0, \mathsf{K}(\mathsf{x}, \mathsf{x}'))$$

$$\mathsf{K}(\mathsf{x}, \mathsf{x}') = \sigma^2 \exp\left(-\frac{1}{2l^2}(\mathsf{x} - \mathsf{x}')^2\right)$$ (smoothly wiggling functions expected)

sum of Gaussian variables = Gaussian: induces a GP over $\mathsf{y}(\mathsf{x})$

$$p(\mathsf{y}(\mathsf{x})|\theta) = \mathcal{GP}(0, \mathsf{K}(\mathsf{x}, \mathsf{x}') + \mathrm{I}\sigma_{\mathsf{y}}^2)$$

Mathematical Foundations: Marginalization

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

Mathematical Foundations: Marginalization

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) \mathrm{d}\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \right)$$

Mathematical Foundations: Marginalization

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) \mathrm{d}\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B}^\mathrm{T} & \mathrm{C} \end{bmatrix}\right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{a}, \mathrm{A})$$

Mathematical Foundations: Marginalization

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

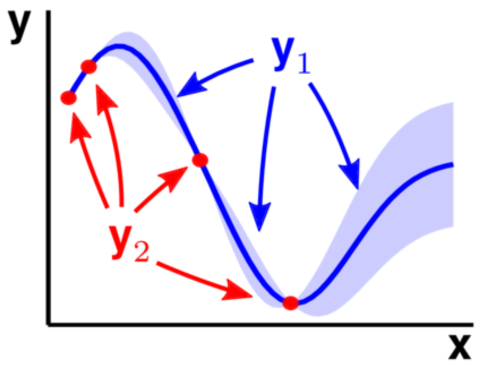
We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) d\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}\right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{a}, A)$$

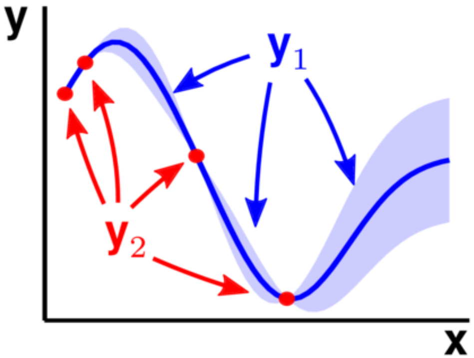
$\implies$ Only need to represent finite dimensional projections of GPs on computer.

Q3: How do we Make Predictions?



Q3: How do we Make Predictions?

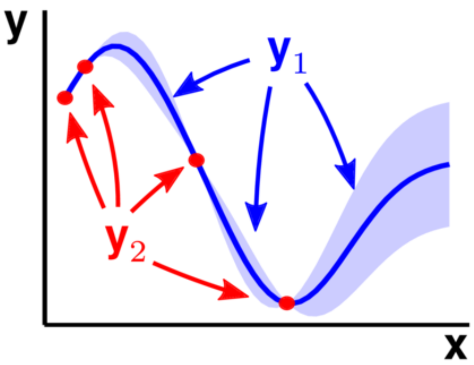
$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^{\mathsf{T}} & \mathsf{C} \end{bmatrix} \right)$$



Q3: How do we Make Predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}\right)$$

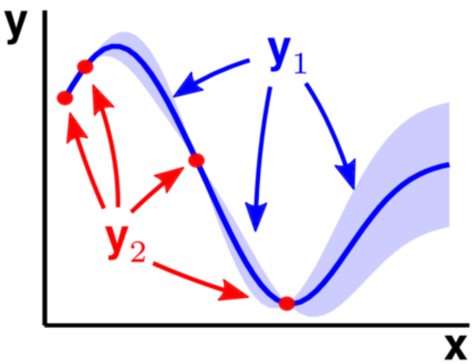
$$p(\mathbf{y}_1 | \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$



Q3: How do we Make Predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\mathsf{T} & \mathsf{C} \end{bmatrix}\right)$$

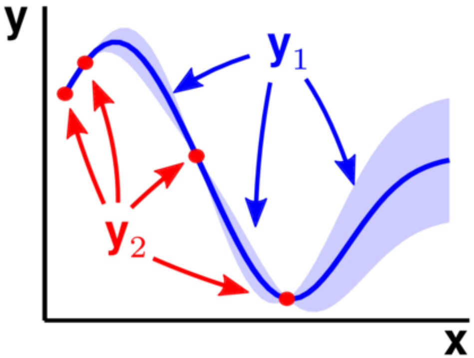
$$p(\mathbf{y}_1 | \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$



$$\implies p(\mathbf{y}_1 | \mathbf{y}_2) = \mathcal{N}(\mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b}), \mathsf{A} - \mathsf{B}\mathsf{C}^{-1}\mathsf{B}^\mathsf{T})$$

Q3: How do we Make Predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\mathsf{T} & \mathsf{C} \end{bmatrix}\right)$$



$$p(\mathbf{y}_1 | \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$

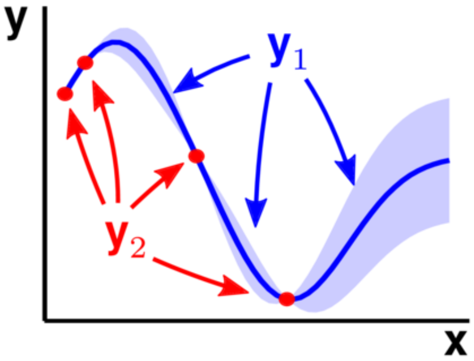
$$\Longrightarrow \quad p(\mathbf{y}_1 | \mathbf{y}_2) = \mathcal{N}(\underline{\mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b})}, \mathsf{A} - \mathsf{B}\mathsf{C}^{-1}\mathsf{B}^\mathsf{T})$$

predictive mean

$$\mu_{\mathbf{y}_1|\mathbf{y}_2} = \mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b})$$

Q3: How do we Make Predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\mathsf{T} & \mathsf{C} \end{bmatrix}\right)$$



$$p(\mathbf{y}_1 | \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$

$$\implies p(\mathbf{y}_1 | \mathbf{y}_2) = \mathcal{N}(\underline{\mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b})}, \mathsf{A} - \mathsf{B}\mathsf{C}^{-1}\mathsf{B}^\mathsf{T})$$

predictive mean

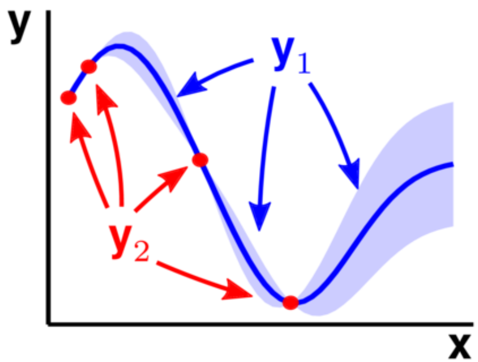
$$\begin{aligned} \mu_{\mathbf{y}_1|\mathbf{y}_2} &= \mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b}) \\ &= \mathsf{B}\mathsf{C}^{-1}\mathbf{y}_2 \\ &= \mathsf{W}\mathbf{y}_2 \end{aligned}$$

linear in the data

Q3: How do we Make Predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\top & \mathsf{C} \end{bmatrix}\right)$$

$$p(\mathbf{y}_1|\mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$



$$\Longrightarrow \quad p(\mathbf{y}_1|\mathbf{y}_2) = \mathcal{N}(\underline{\mathbf{a} + \mathsf{BC}^{-1}(\mathbf{y}_2 - \mathbf{b})}, \underline{\mathsf{A} - \mathsf{BC}^{-1}\mathsf{B}^\top})$$

predictive mean

$$\mu_{\mathbf{y}_1|\mathbf{y}_2} = \mathbf{a} + \mathsf{BC}^{-1}(\mathbf{y}_2 - \mathbf{b})$$
$$= \mathsf{BC}^{-1}\mathbf{y}_2$$
$$= \mathsf{W}\mathbf{y}_2$$

linear in the data

predictive covariance

$$\Sigma_{\mathbf{y}_1|\mathbf{y}_2} = \mathsf{A} - \mathsf{BC}^{-1}\mathsf{B}^\top$$

predictive uncertainty = prior uncertainty − reduction in uncertainty

predictions more confident than prior

Partitioned Gaussian Distributions

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad N(x|\mu,\Sigma) = \frac{1}{2\pi^{D/2}}\frac{1}{|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

$$\mathbf{x} = \begin{pmatrix}\mathbf{x}_a \\ \mathbf{x}_b\end{pmatrix} \begin{array}{l}\text{Mx1} \\ \text{(D-M)x1}\end{array} \qquad \boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b\end{pmatrix}$$

$$\underset{\text{DxD}}{\boldsymbol{\Sigma}} = \begin{array}{cc} \text{MxM} & \text{Mx(D-M)} \\ \end{array}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix}\boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb}\end{pmatrix}$$

$$\begin{array}{cc} \text{(D-M)xM} & \text{(D-M)x(D-M)} \end{array}$$

precision matrix $\boldsymbol{\Lambda} \equiv \boldsymbol{\Sigma}^{-1}$

$$\boldsymbol{\Lambda} = \begin{pmatrix}\boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb}\end{pmatrix}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix}$$

where $M = (A - BD^{-1}C)^{-1}$

Thus, $\Lambda_{aa} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$

Partitioned Conditional Distributions

- Conditional distribution $p(x_a|x_b)$ can be evaluated from the joint distribution $p(x) = p(x_a, x_b)$ by fixing x_b
- That is $p(x_a|x_b) \propto p(x_a, x_b)$ if we treat x_b as constant
- $\Delta^2 = -\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)$
- $= -\frac{1}{2}\left((x_a-\mu_a)^T \quad (x_b-\mu_b)^T\right)\begin{bmatrix}\Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb}\end{bmatrix}\begin{pmatrix}x_a-\mu_a \\ x_b-\mu_b\end{pmatrix}$
- $= -\frac{1}{2}(x_a-\mu_a)^T\Lambda_{aa}(x_a-\mu_a) - \frac{1}{2}(x_a-\mu_a)^T\Lambda_{ab}(x_b-\mu_b)$
  $- \frac{1}{2}(x_b-\mu_b)^T\Lambda_{ba}(x_a-\mu_a) - \frac{1}{2}(x_b-\mu_b)^T\Lambda_{bb}(x_b-\mu_b)$
- 2nd order of x_a: $-\frac{1}{2}(x_a)^T\Lambda_{aa}(x_a)$
- 1st order of x_a: $x_a^T\Lambda_{aa}\mu_a - \frac{1}{2}x_a^T\Lambda_{ab}(x_b-\mu_b) - \frac{1}{2}x_a^T\Lambda_{ba}^T(x_b-\mu_b)$

Completing the squares of the exponent

- General form $\Delta^2 = -\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)$
  $= -\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x} + \mathbf{x}^T\Sigma^{-1}\boldsymbol{\mu} + const$
- Coefficient of the 2nd order term: covariance matrix $\sum^{-1}$
- Coefficient of the 1st order term: $\sum^{-1}\mu$

2nd order of x_a $\quad -\frac{1}{2}\mathbf{x}_a^T\Lambda_{aa}\mathbf{x}_a$ $\quad \Rightarrow \Sigma_{a|b} = \Lambda_{aa}^{-1}$

1st order of x_a $\quad \mathbf{x}_a^T\Lambda_{aa}\boldsymbol{\mu}_a - \frac{1}{2}\mathbf{x}_a^T\Lambda_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b) - \frac{1}{2}\mathbf{x}_a^T\Lambda_{ba}^T(\mathbf{x}_b - \boldsymbol{\mu}_b)$

$$= x_a^T\{\Lambda_{aa}\mu_a - \Lambda_{ab}(x_b - \mu_b)\}$$
$$= x_a^T\Sigma_{a|b}^{-1}\mu_{a|b}$$

$$\Rightarrow \mu_{a|b} = \Sigma_{a|b}\{\Lambda_{aa}\mu_a - \Lambda_{ab}(x_b - \mu_b)\} = \mu_a - \Lambda_{aa}^{-1}\Lambda_{ab}(x_b - \mu_b)$$

Conditional Distribution of Multivariate Gaussian

- $\Sigma_{a|b} = \Lambda_{\mathrm{aa}}^{-1} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$
- $\mu_{a|b} = \mu_a - \Lambda_{aa}^{-1}\Lambda_{ab}(x_b - \mu_b)$
- Moreover, $\begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$
- and, $\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix}$,
- $M = (A - BD^{-1}C)^{-1}$
- Thus, $\Lambda_{aa} = \left(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}\right)^{-1}$,
- $\Lambda_{ab} = -\left(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}\right)^{-1}\Sigma_{ab}\Sigma_{bb}^{-1}$
- As a result, $\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$

Conditional Distribution of Multivariate Gaussian

- Let $x = [x_1\ x_2 \cdots x_D]$ be jointly Gaussian
- $x \sim MN(\mu, \Sigma)$
- Partitioned x into two groups $x = \begin{bmatrix} x_a \\ x_b \end{bmatrix}$; $x_a = [x_1\ x_2\ \ldots x_M]^T$, and $x_b = [x_{M+1}\ \ldots x_D]^T$
- Think of x_a as the target variables and x_b as the features.
- $p(x_a|x_b) = MN(\mu_{a|b}, \Sigma_{a|b})$
- $\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$
- $\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix} \begin{array}{l} \text{Mx1} \\ \text{(D-M)x1} \end{array} \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix}$$

$$\underset{\text{DxD}}{\boldsymbol{\Sigma}} = \begin{array}{cc} \text{MxM} & \text{Mx(D-M)} \end{array} \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix} \begin{array}{cc} \text{(D-M)xM} & \text{(D-M)x(D-M)} \end{array}$$

Kernel Hyper-parameters

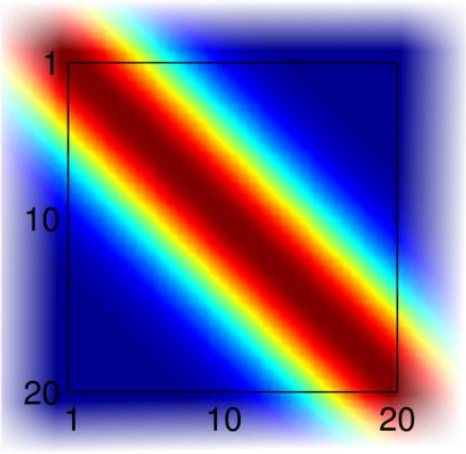
Non-parametric (∞-parametric)

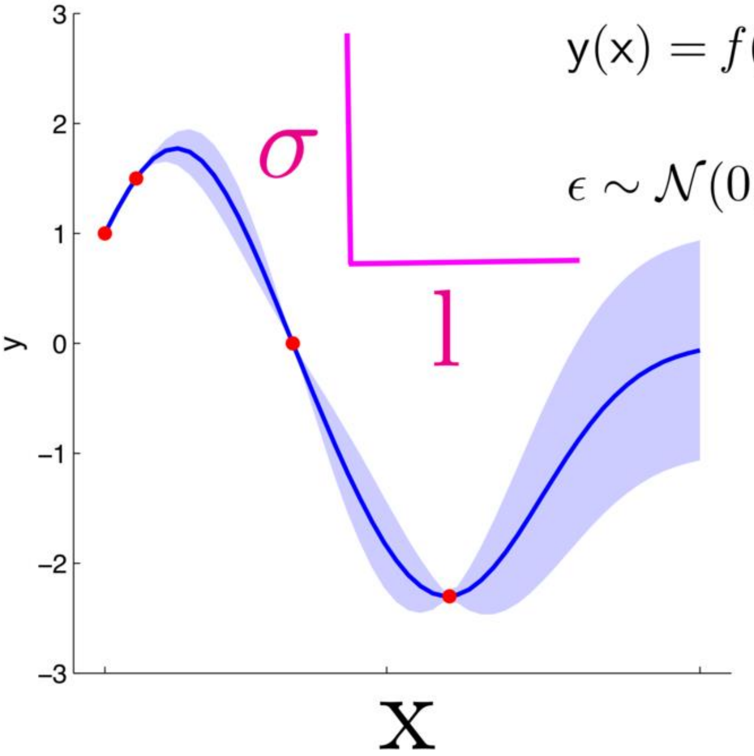
$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

vertical-scale horizontal-scale

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Comparing Three Kernel Parameter Settings

- Changing l will affect the horizontal length scale
- l larger→ correlation decays faster

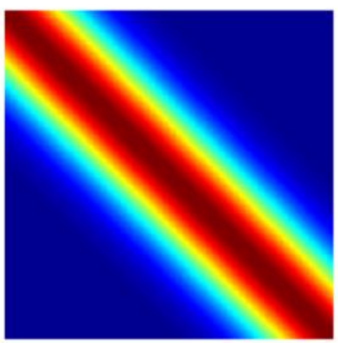
Non-parametric (∞-parametric)

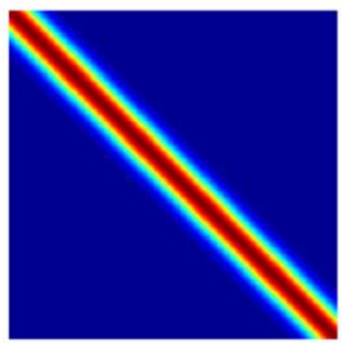
$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

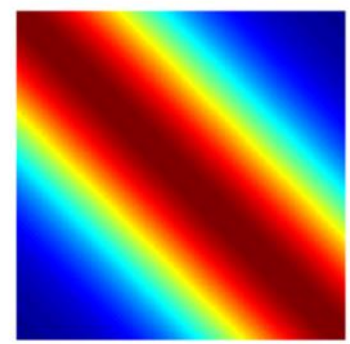
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

- l median | l large | l small
- median scale | short scale | long scale

$\Sigma =$ 

$\Sigma =$ 

$\Sigma =$ 

What effect do the Hyper-parameters Have?

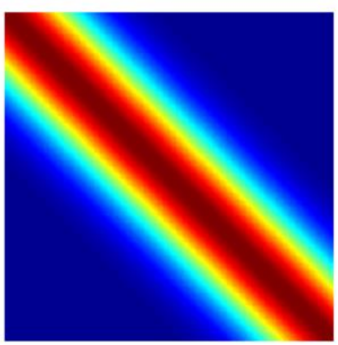
Non-parametric (∞-parametric)

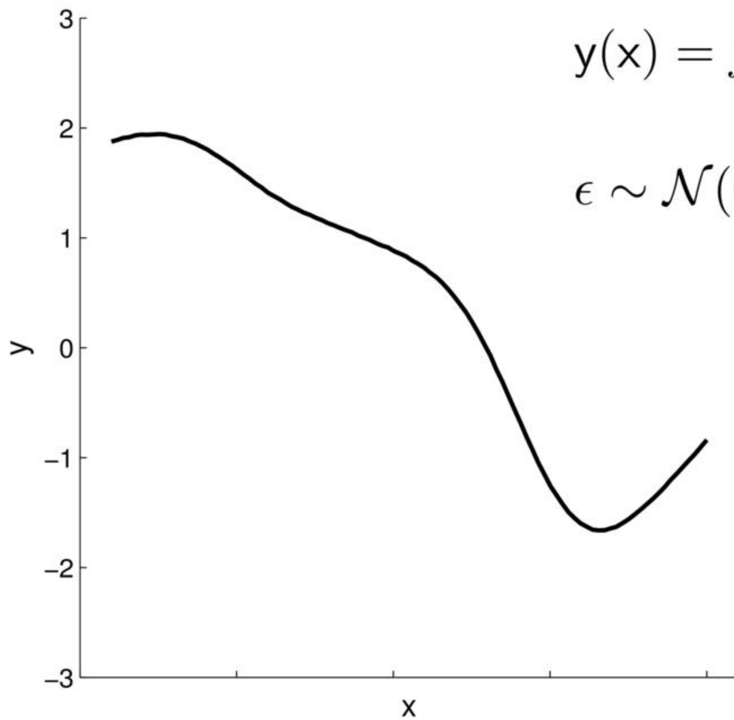
$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$$\Sigma =$$





Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

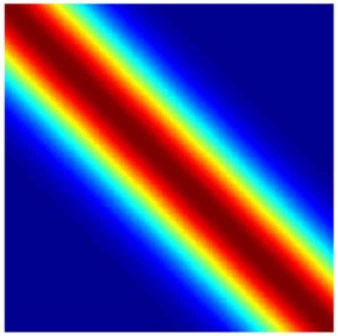
What effect do the Hyper-parameters Have?

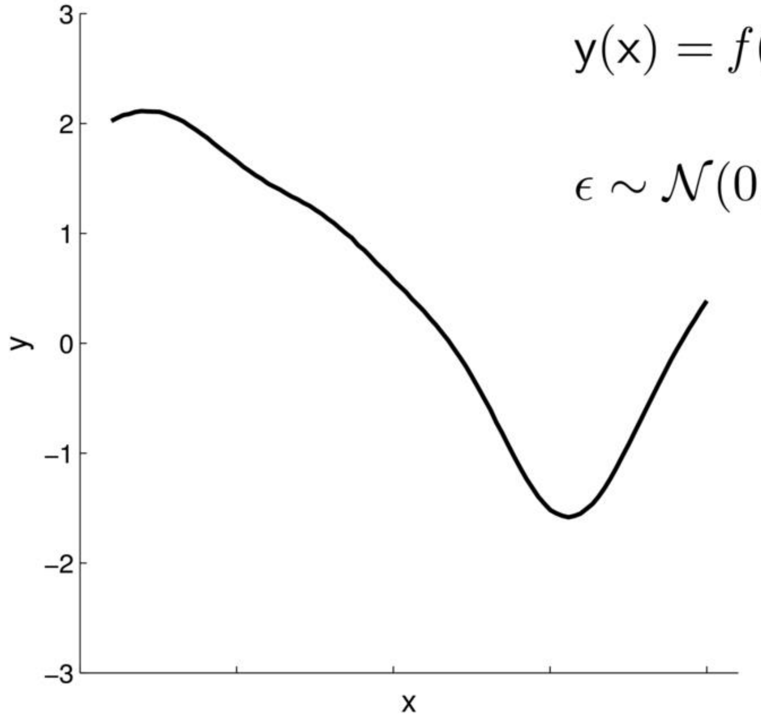
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

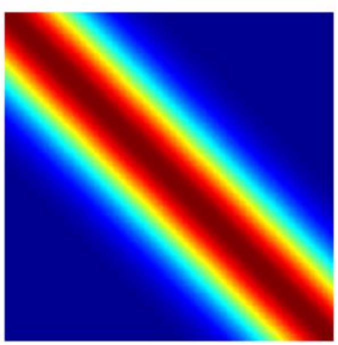
What effect do the Hyper-parameters Have?

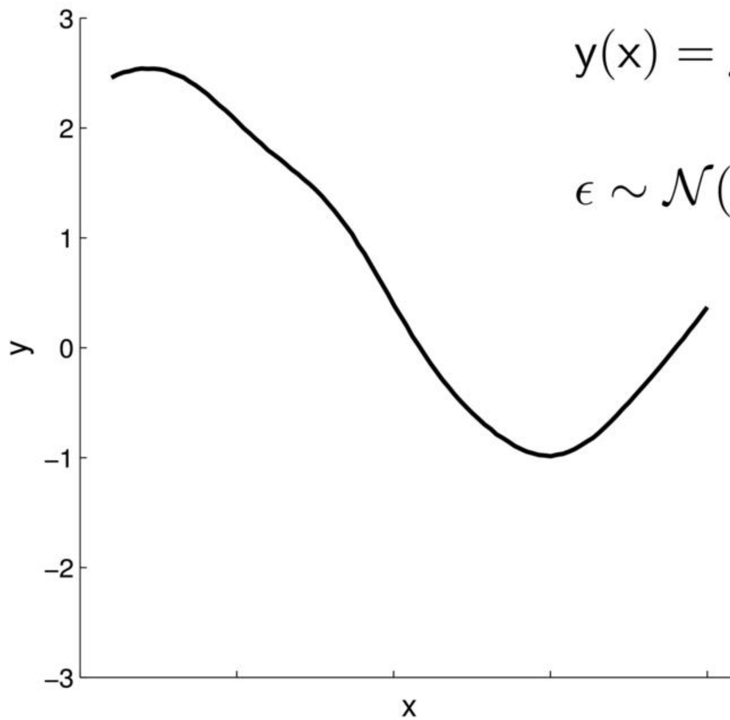
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

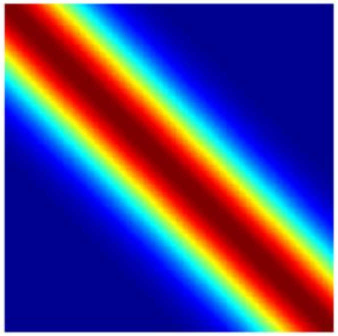
What effect do the Hyper-parameters Have?

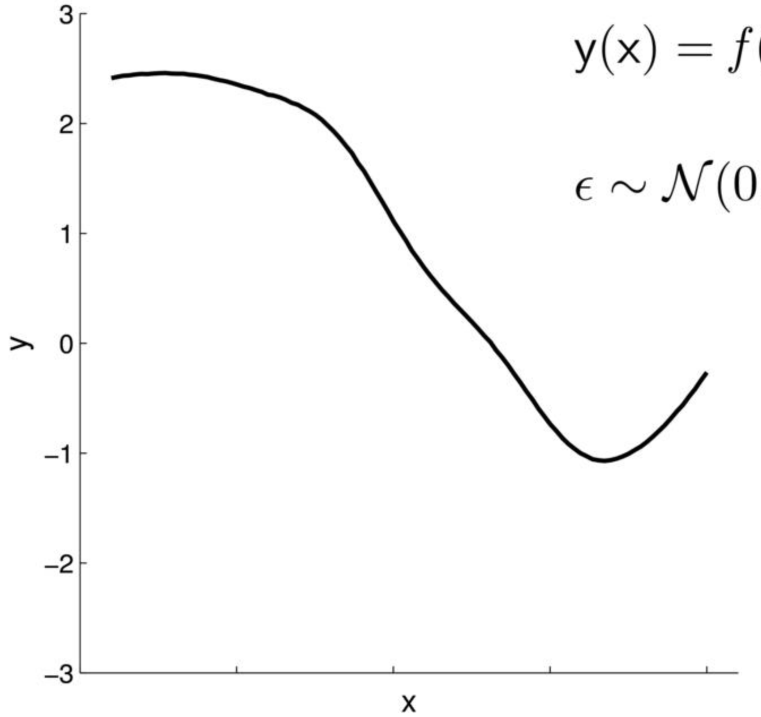
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

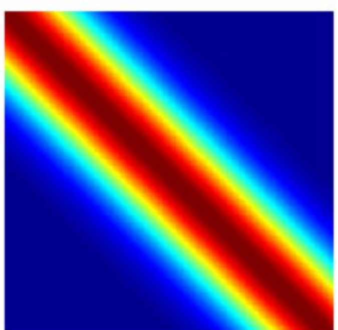
What effect do the Hyper-parameters Have?

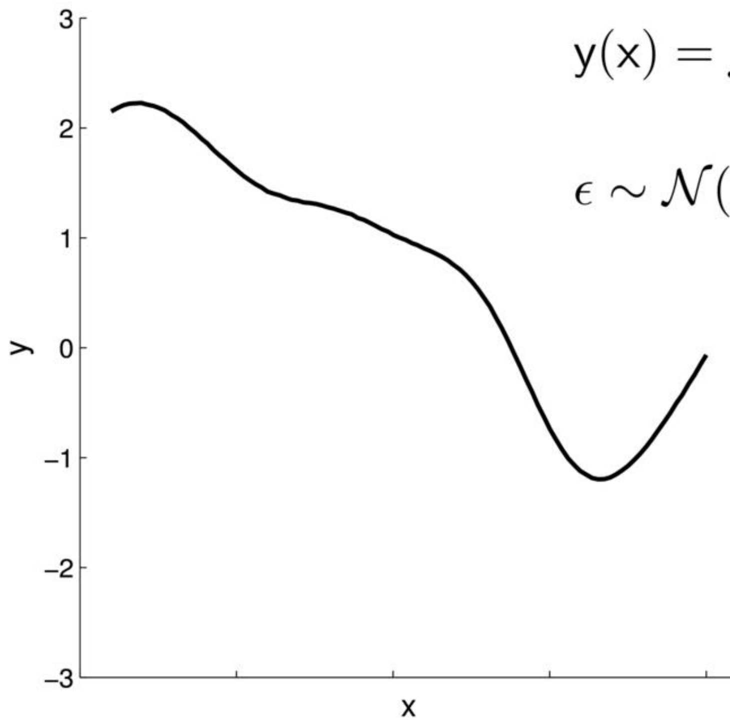
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

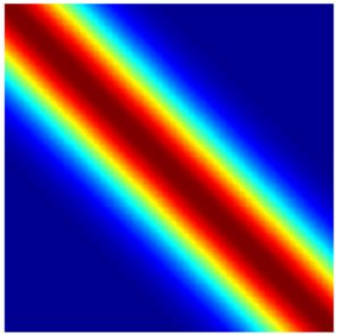
What effect do the Hyper-parameters Have?

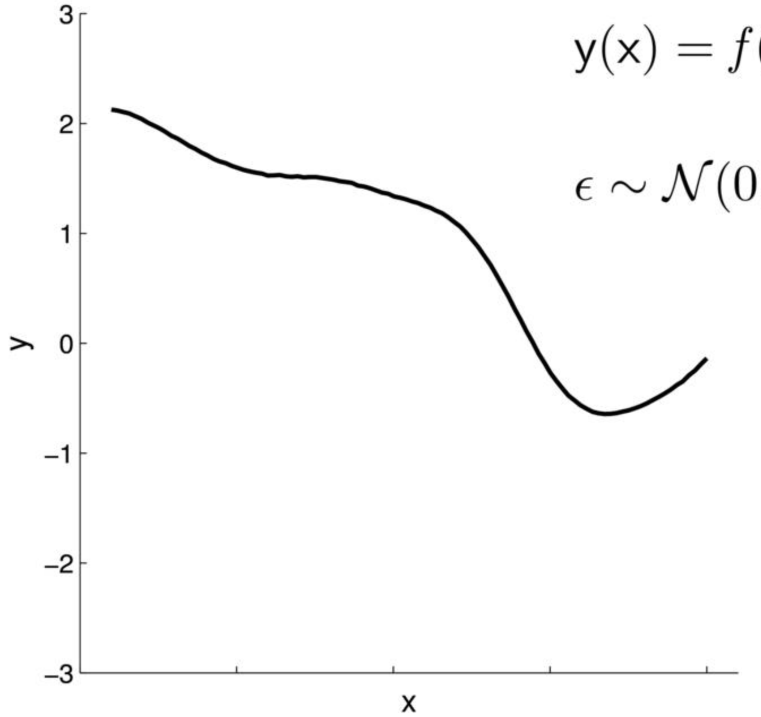
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

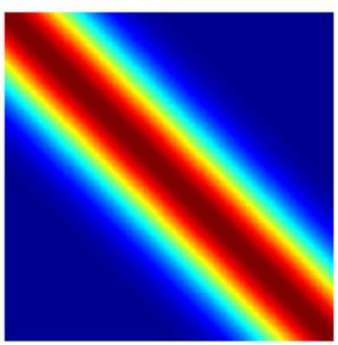
What effect do the Hyper-parameters Have?

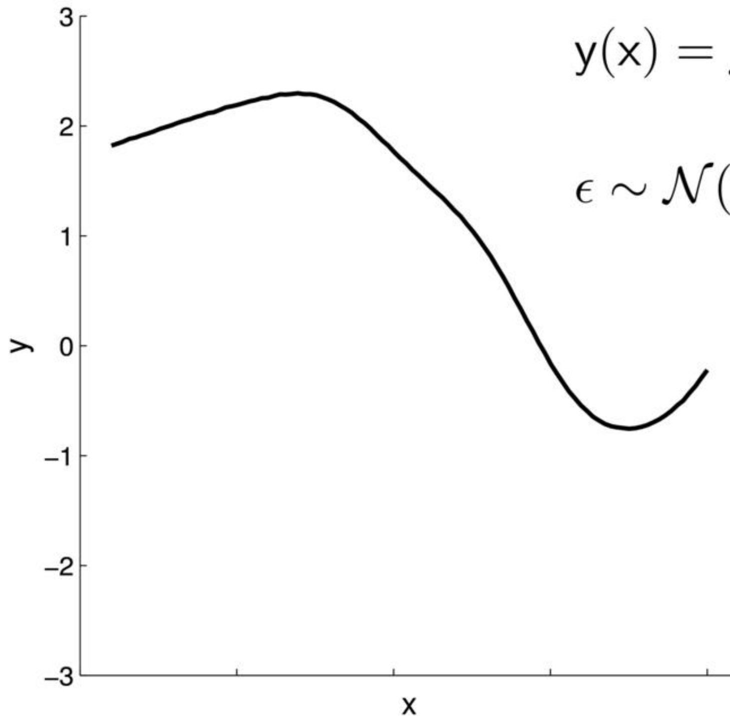
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

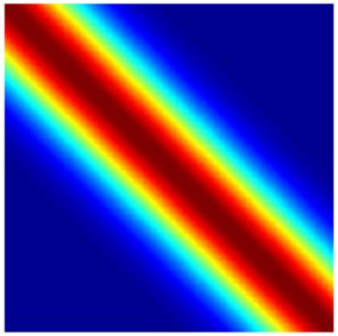
What effect do the Hyper-parameters Have?

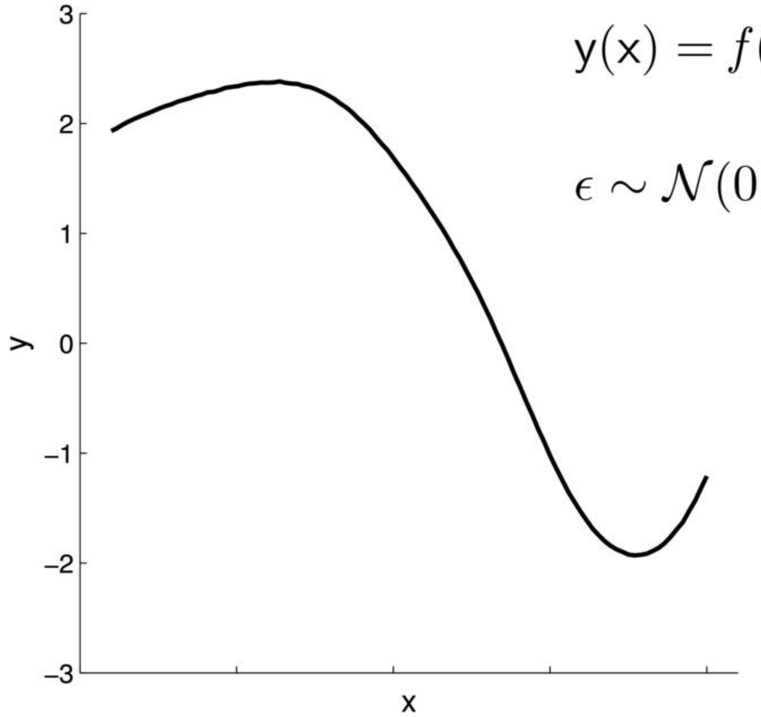
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

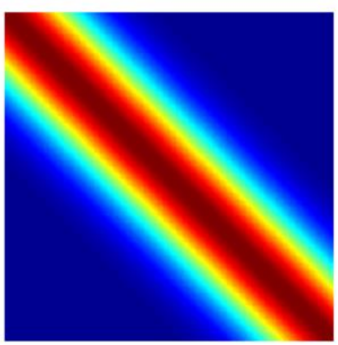
What effect do the Hyper-parameters Have?

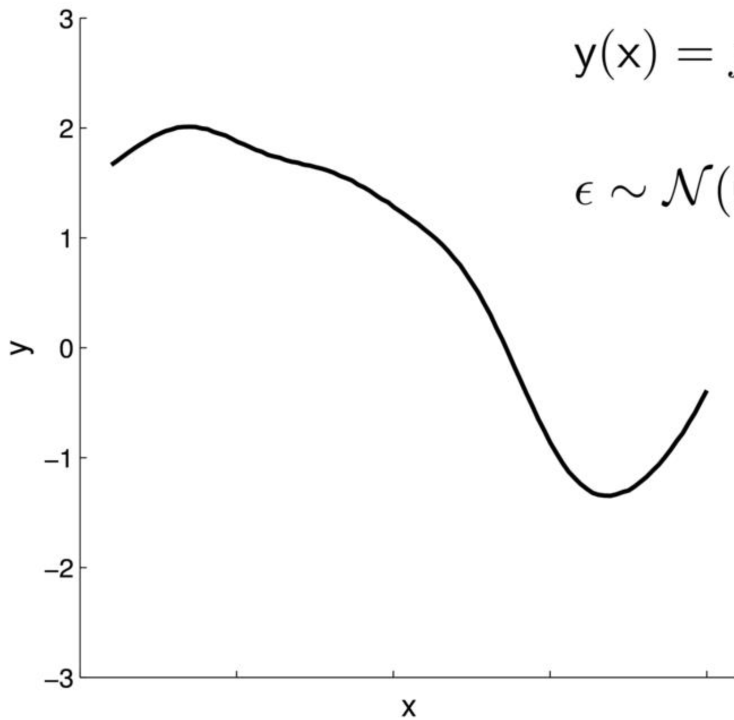
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

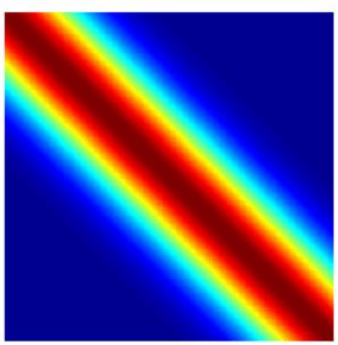
What effect do the Hyper-parameters Have?

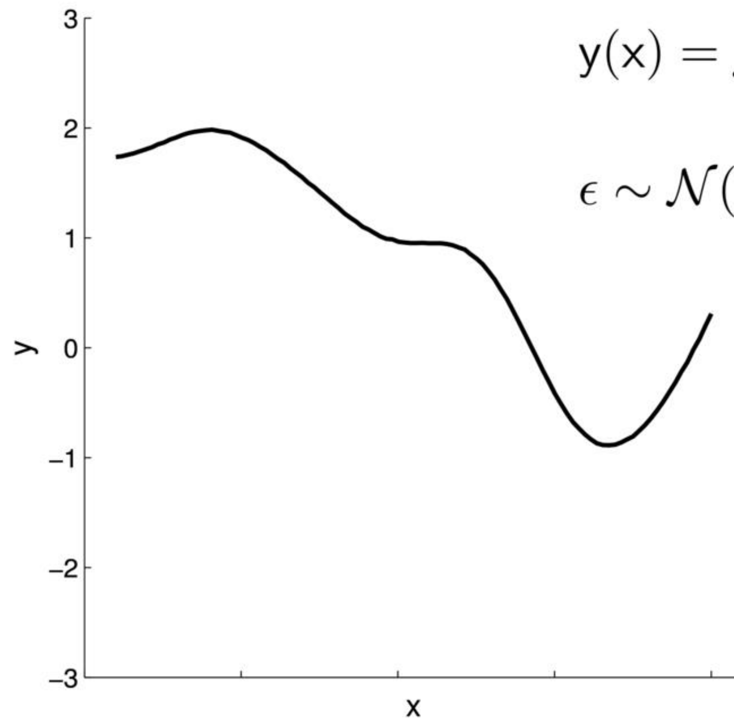
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

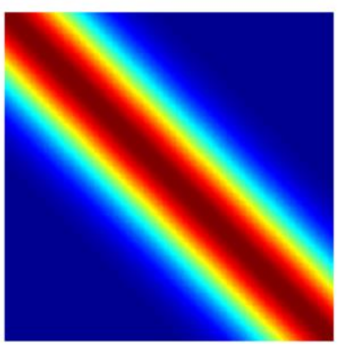
What effect do the Hyper-parameters Have?

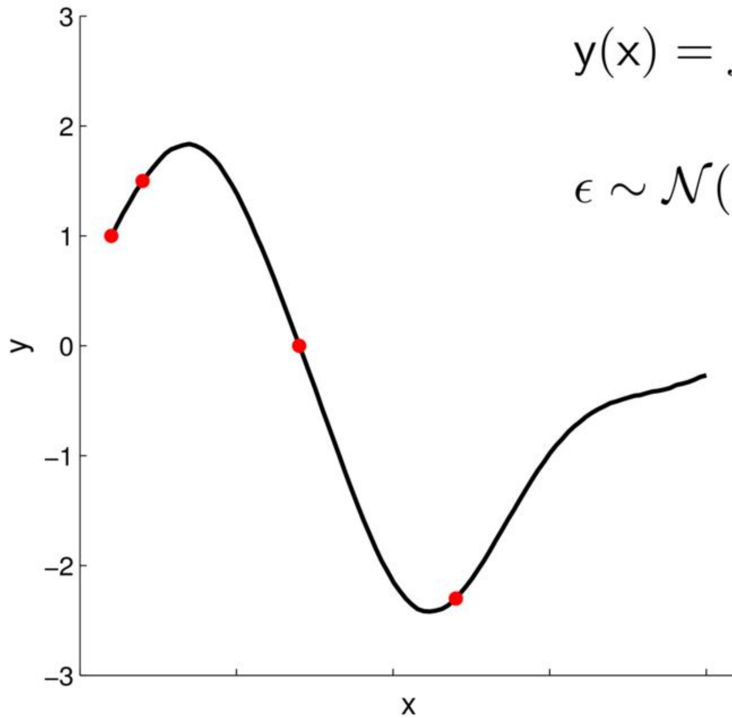
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

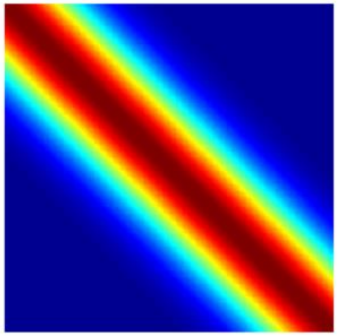
What effect do the Hyper-parameters Have?

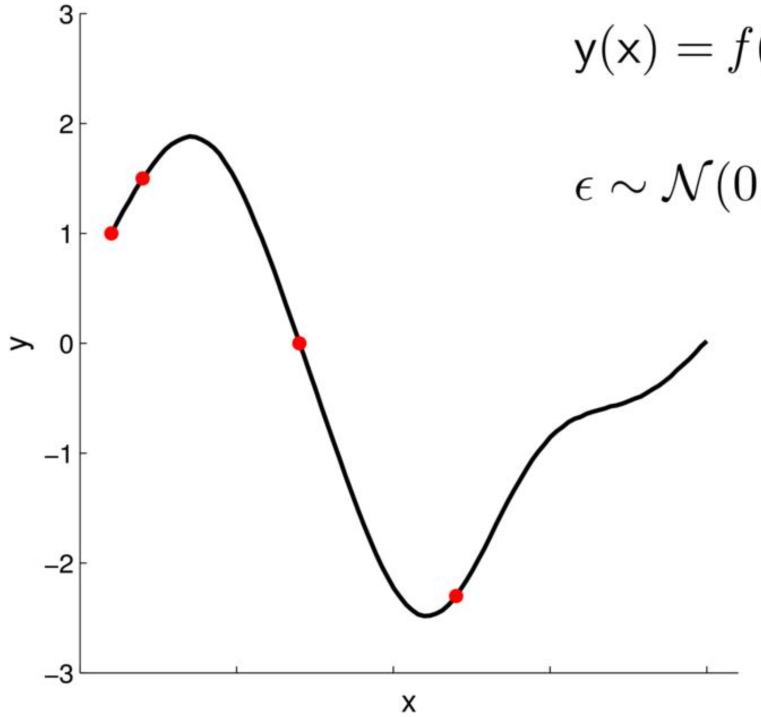
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

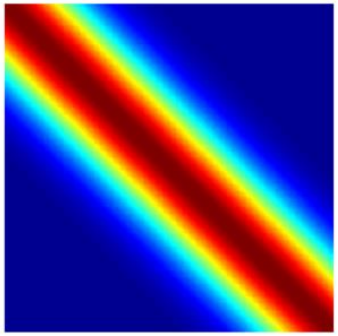
What effect do the Hyper-parameters Have?

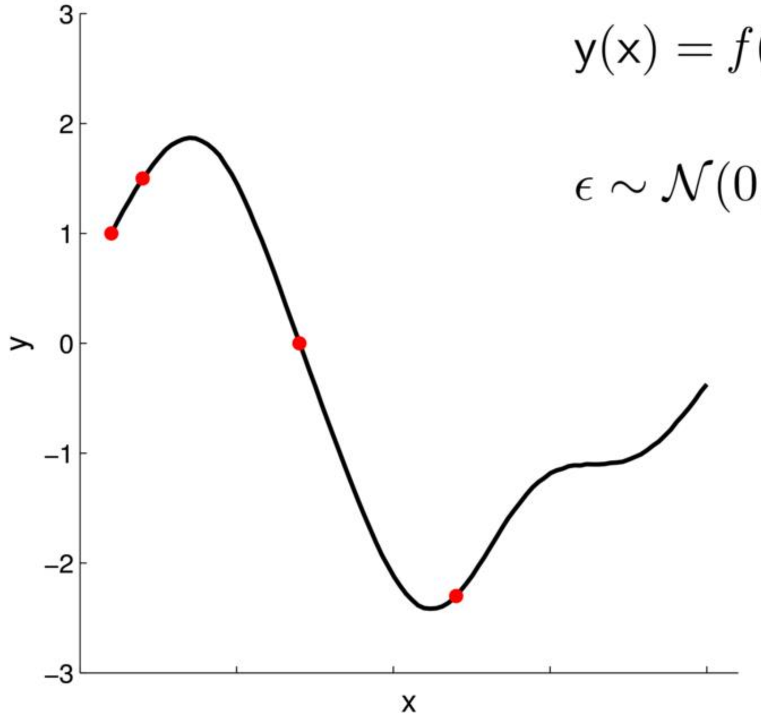
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

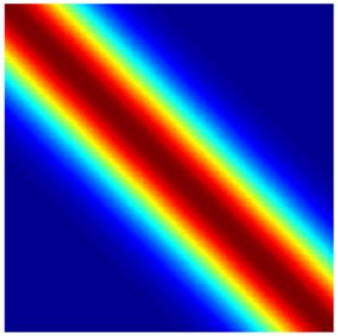
What effect do the Hyper-parameters Have?

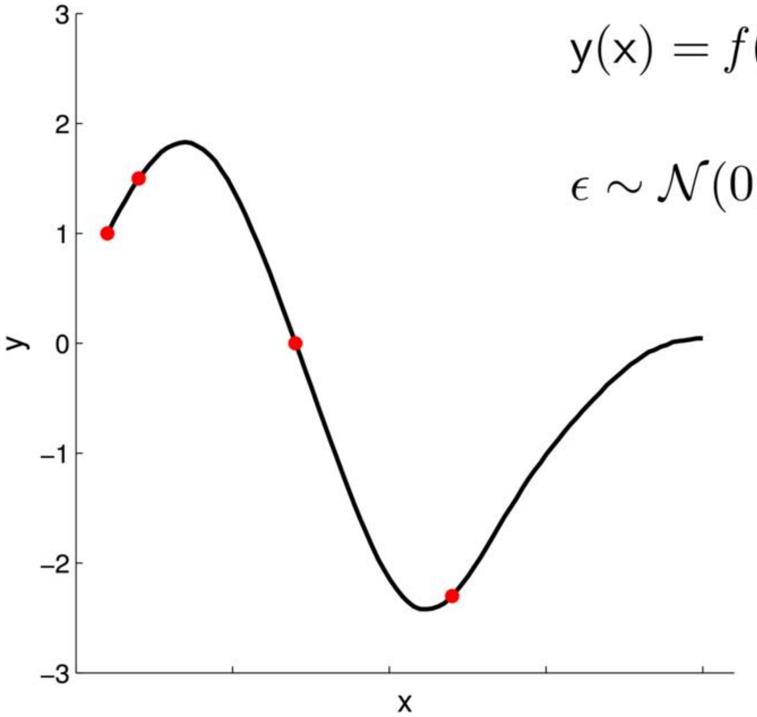
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

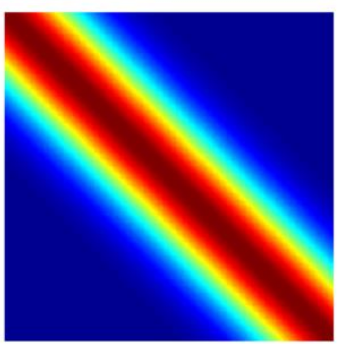
What effect do the Hyper-parameters Have?

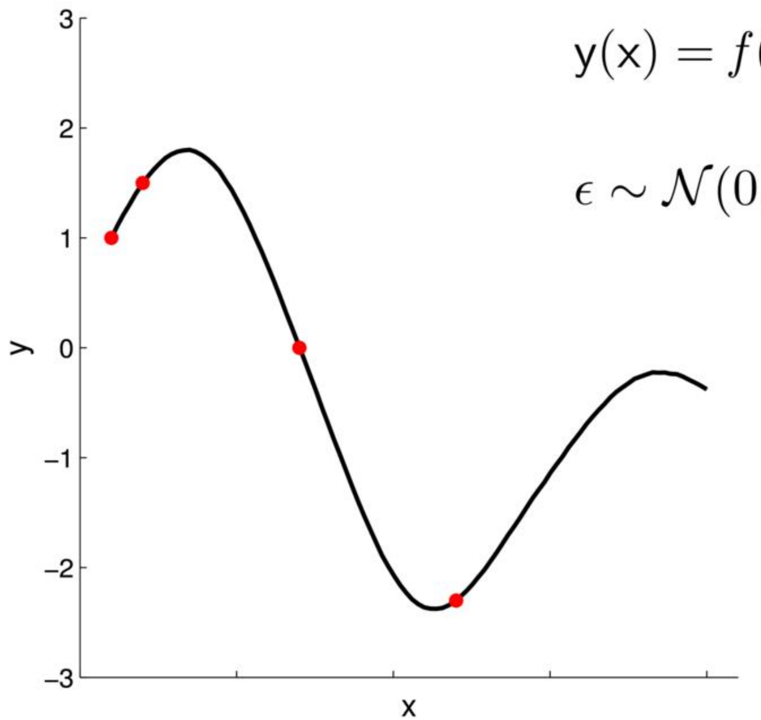
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

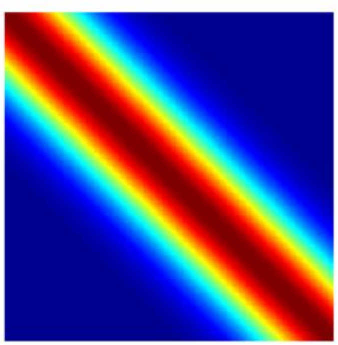
What effect do the Hyper-parameters Have?

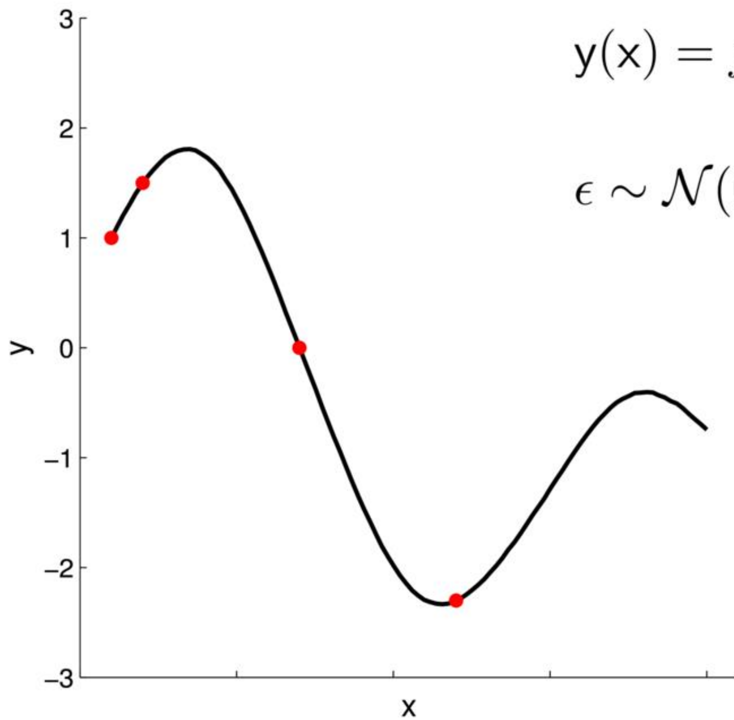
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

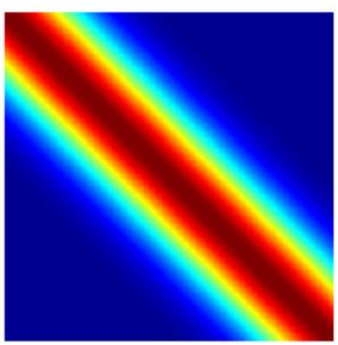
What effect do the Hyper-parameters Have?

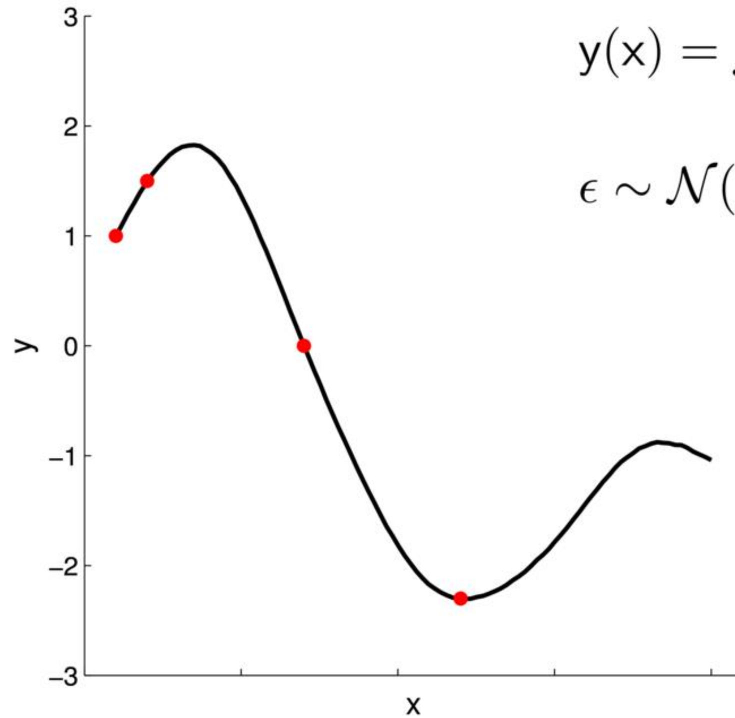
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

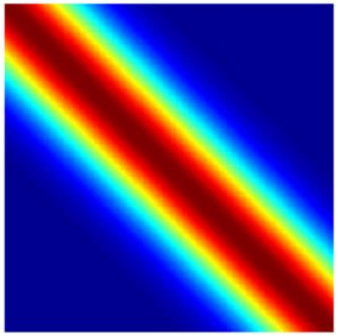
What effect do the Hyper-parameters Have?

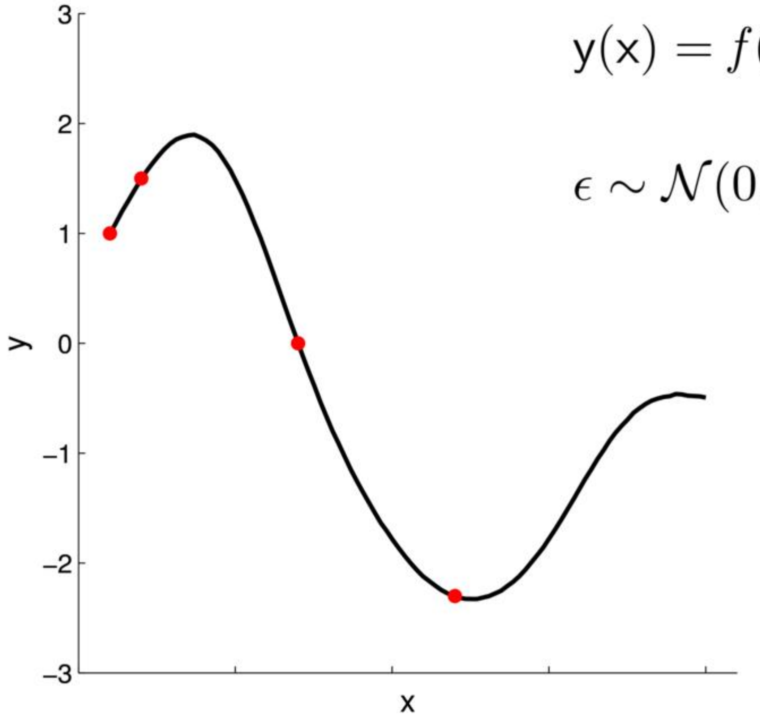
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

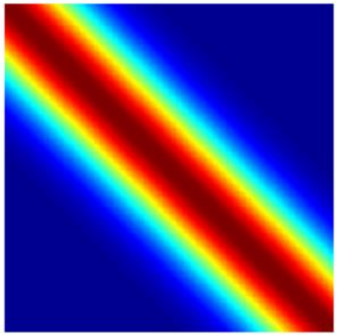
What effect do the Hyper-parameters Have?

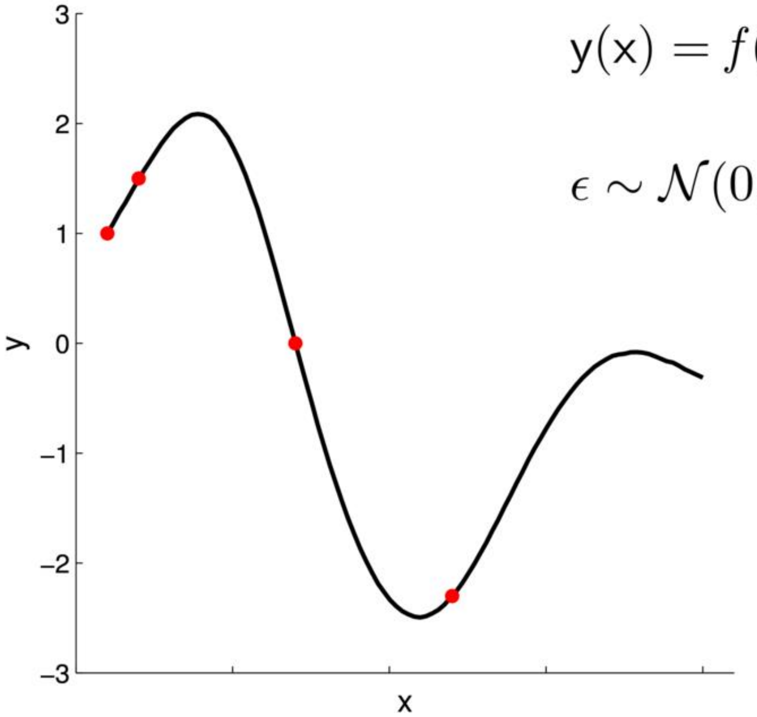
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

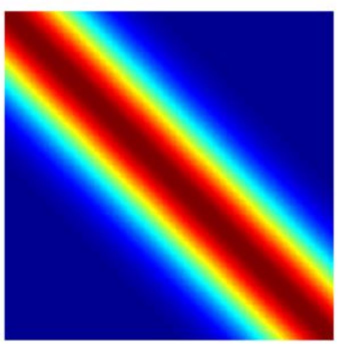
What effect do the Hyper-parameters Have?

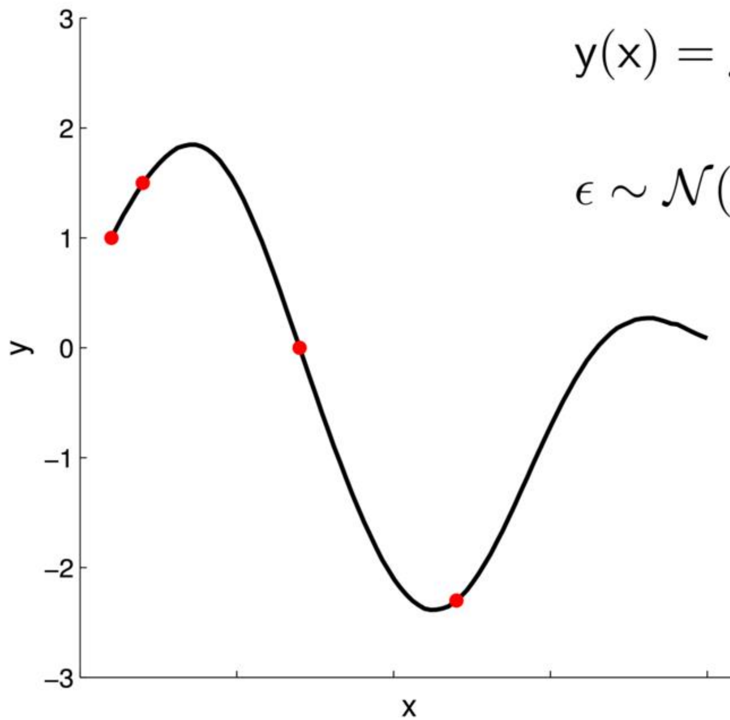
Non-parametric (∞-parametric)

$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$

$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$

$\Sigma =$ 



Parametric model

$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

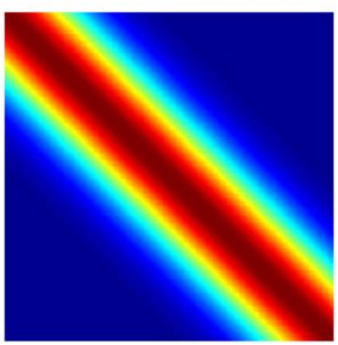
What effect do the Hyper-parameters Have?

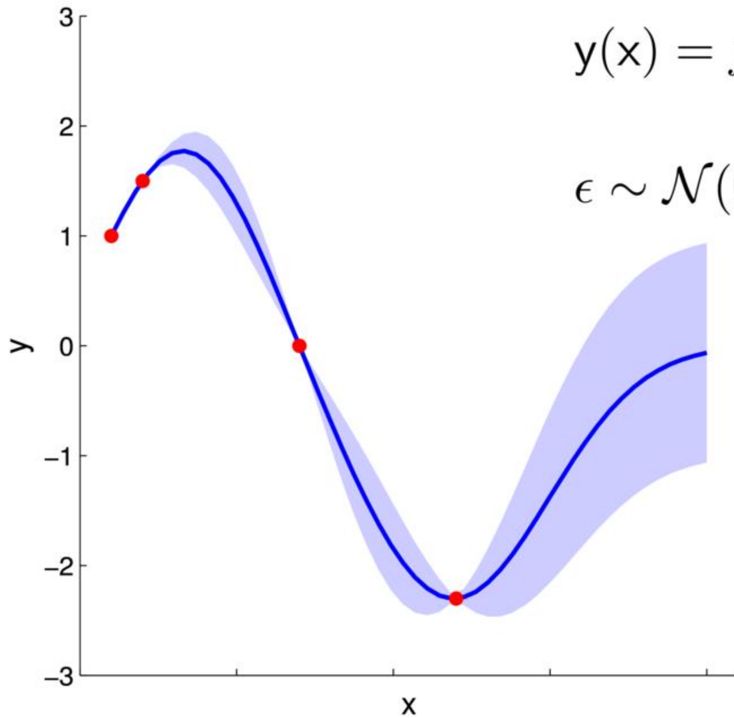
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

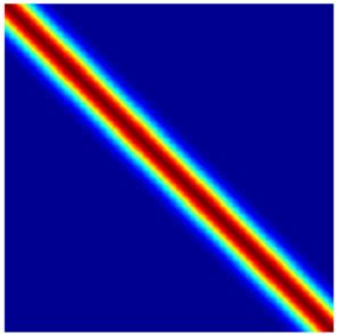
What effect do the Hyper-parameters Have?

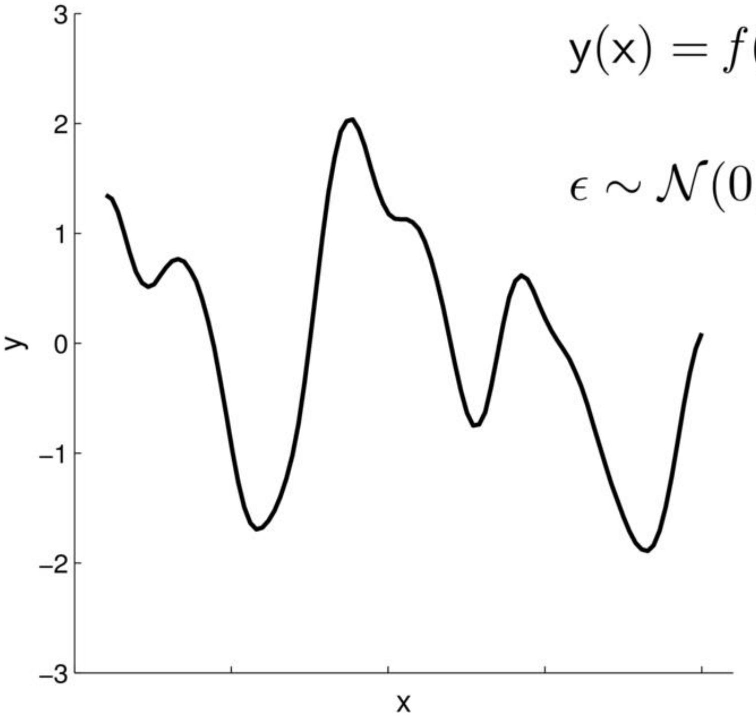
Non-parametric (∞-parametric)

$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$

$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$

$\Sigma =$ 



Parametric model

$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

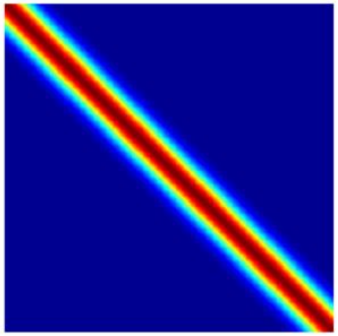
What effect do the Hyper-parameters Have?

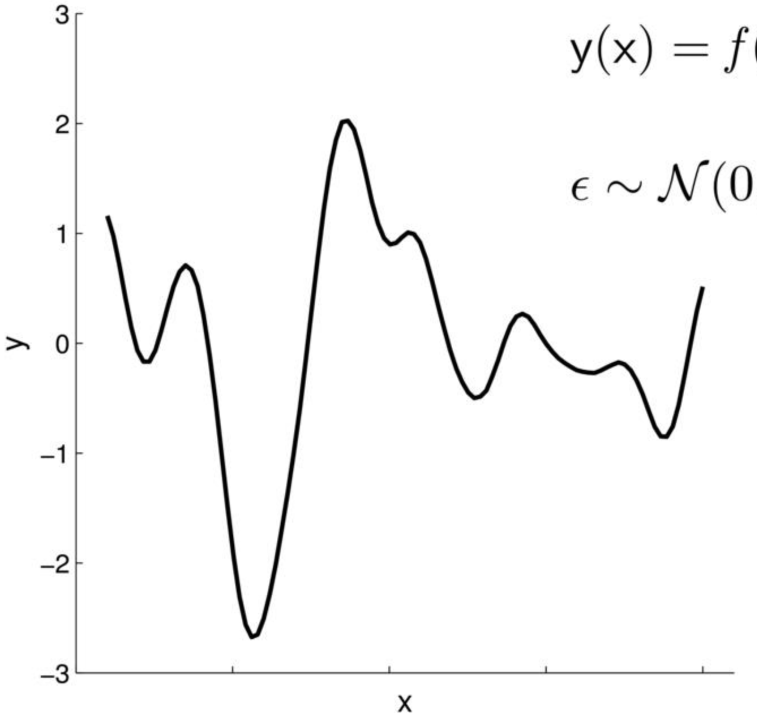
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

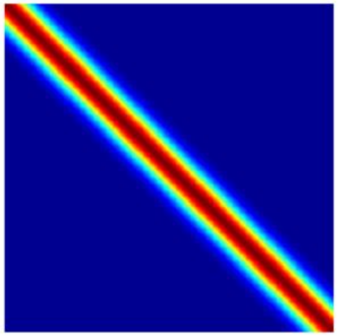
What effect do the Hyper-parameters Have?

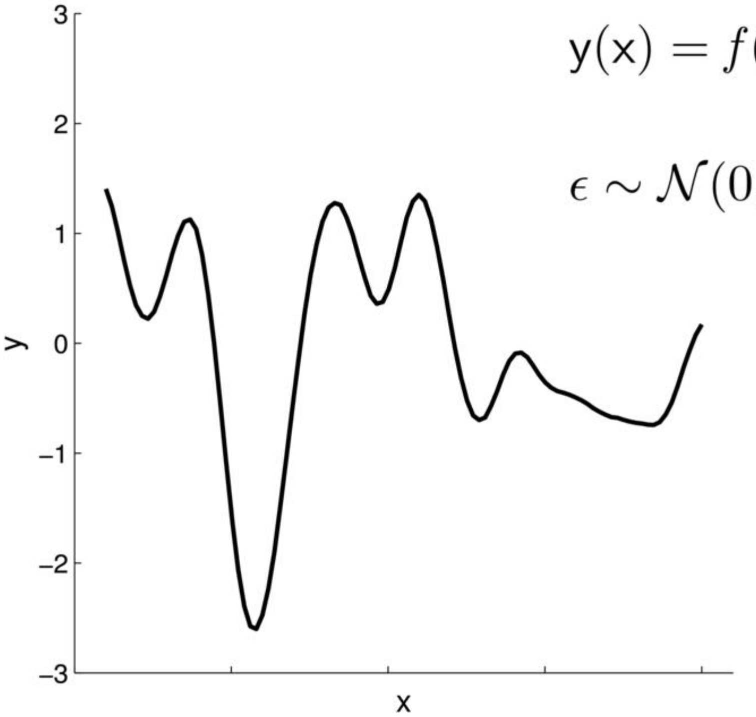
Non-parametric (∞-parametric)

$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$

$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$

$\Sigma =$ 



Parametric model

$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

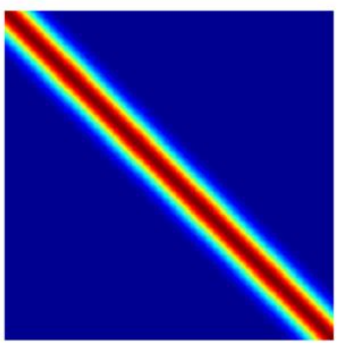
What effect do the Hyper-parameters Have?

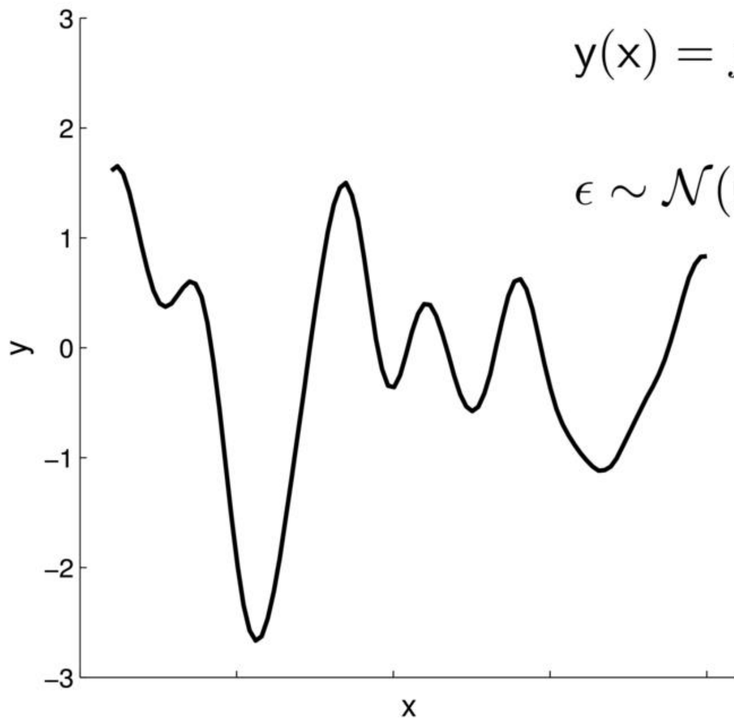
Non-parametric (∞-parametric)

$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$

$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$

$\Sigma =$ 



Parametric model

$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

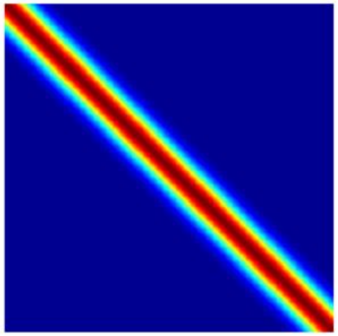
What effect do the Hyper-parameters Have?

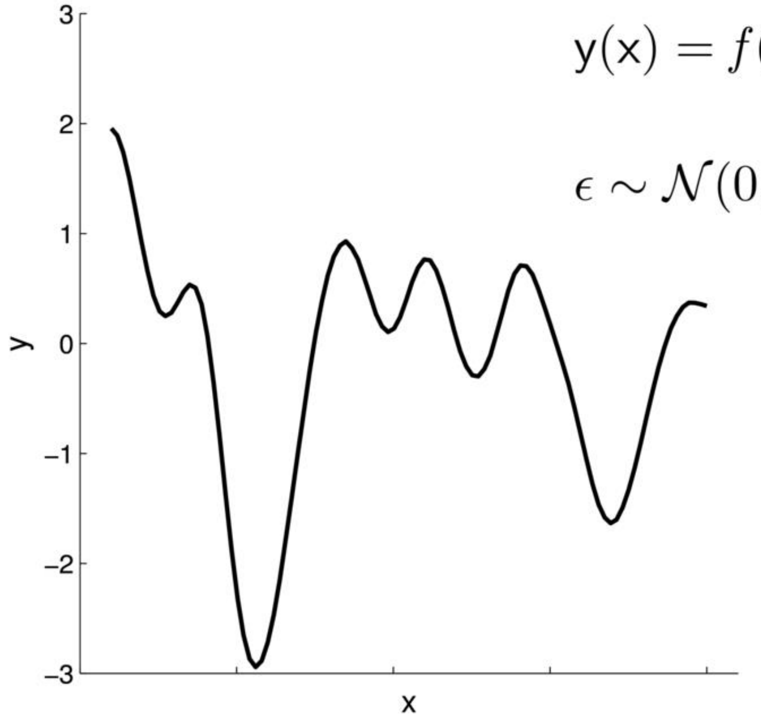
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

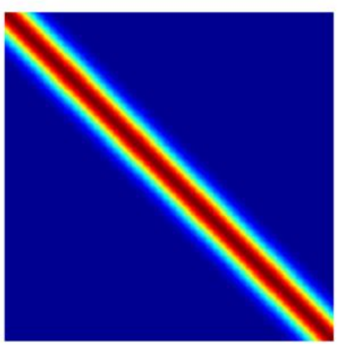
What effect do the Hyper-parameters Have?

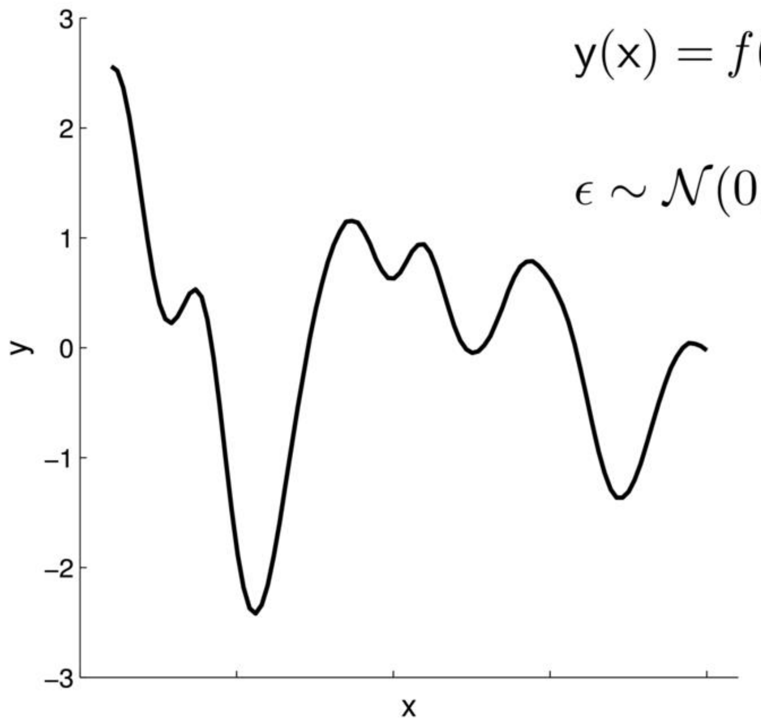
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

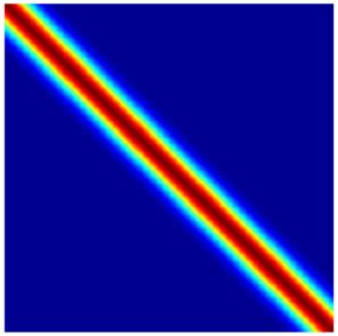
What effect do the Hyper-parameters Have?

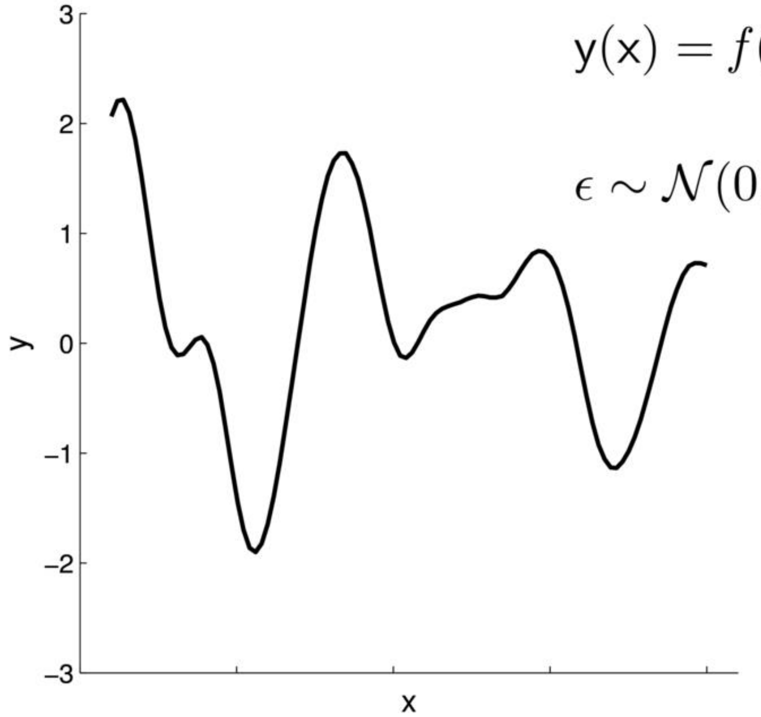
Non-parametric (∞-parametric)

$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$

$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$

$\Sigma =$ 



Parametric model

$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

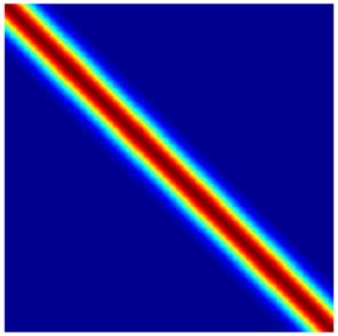
What effect do the Hyper-parameters Have?

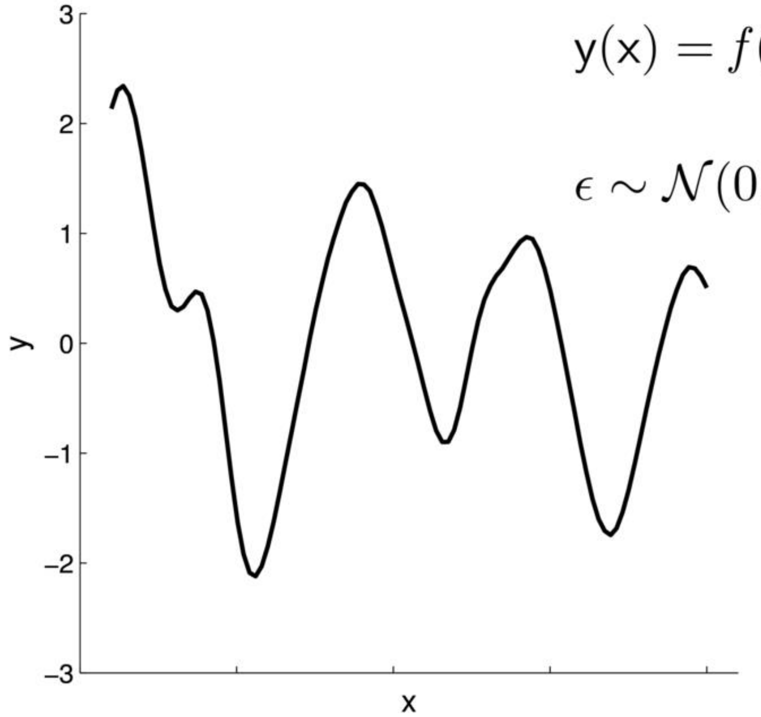
Non-parametric (∞-parametric)

$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$

$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$

$\Sigma =$ 



Parametric model

$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

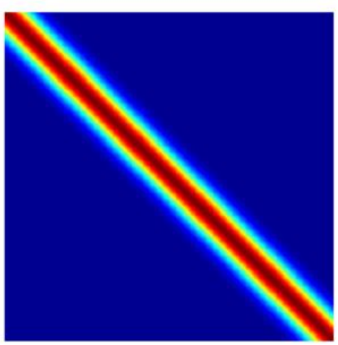
What effect do the Hyper-parameters Have?

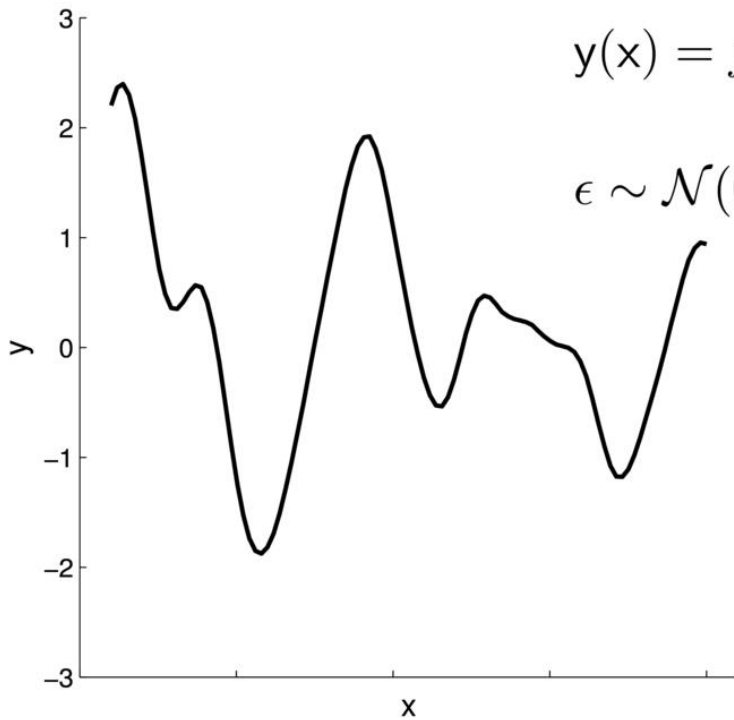
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

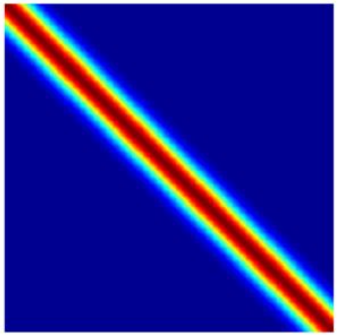
What effect do the Hyper-parameters Have?

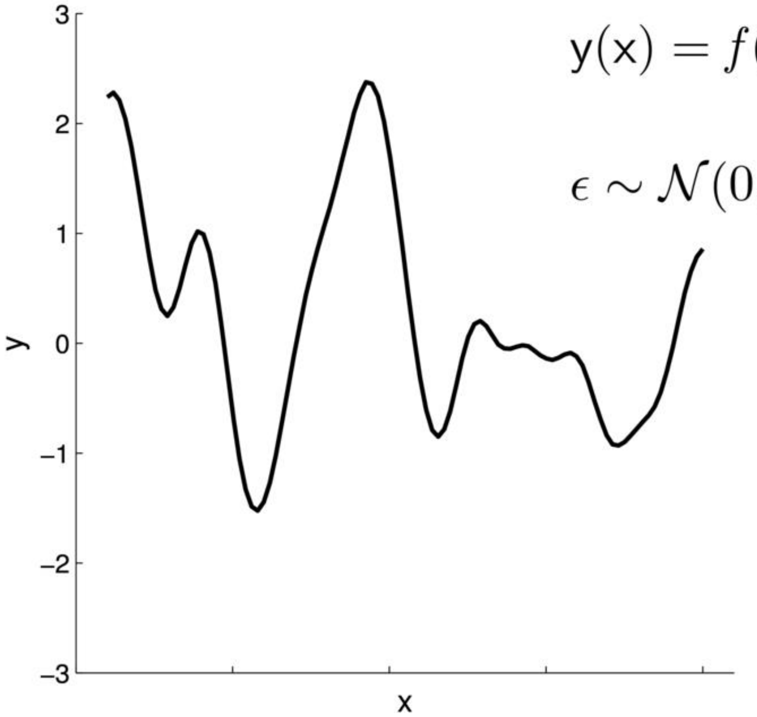
Non-parametric (∞-parametric)

$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$

$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$

$\Sigma =$ 



Parametric model

$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

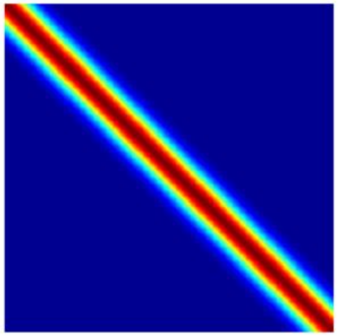
What effect do the Hyper-parameters Have?

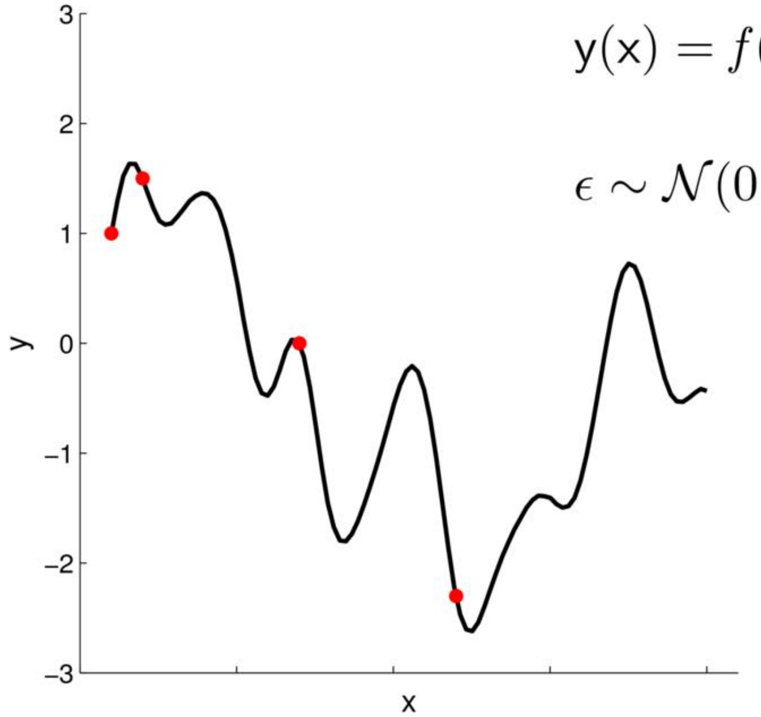
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

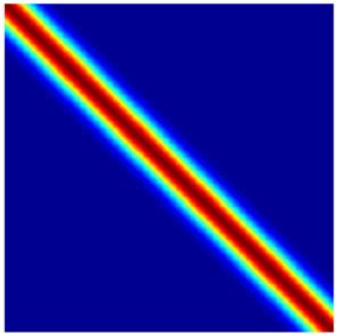
What effect do the Hyper-parameters Have?

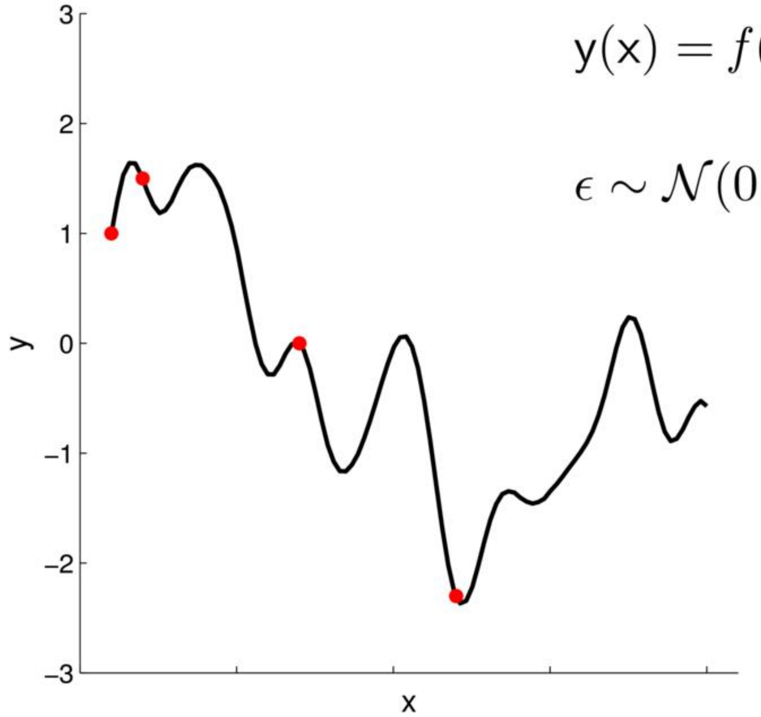
Non-parametric (∞-parametric)

$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$

$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$

$\Sigma =$ 



Parametric model

$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

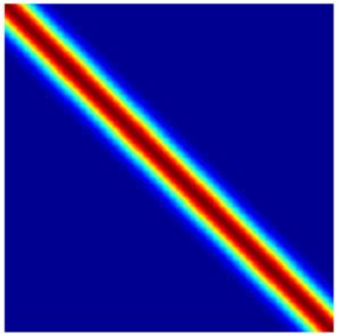
What effect do the Hyper-parameters Have?

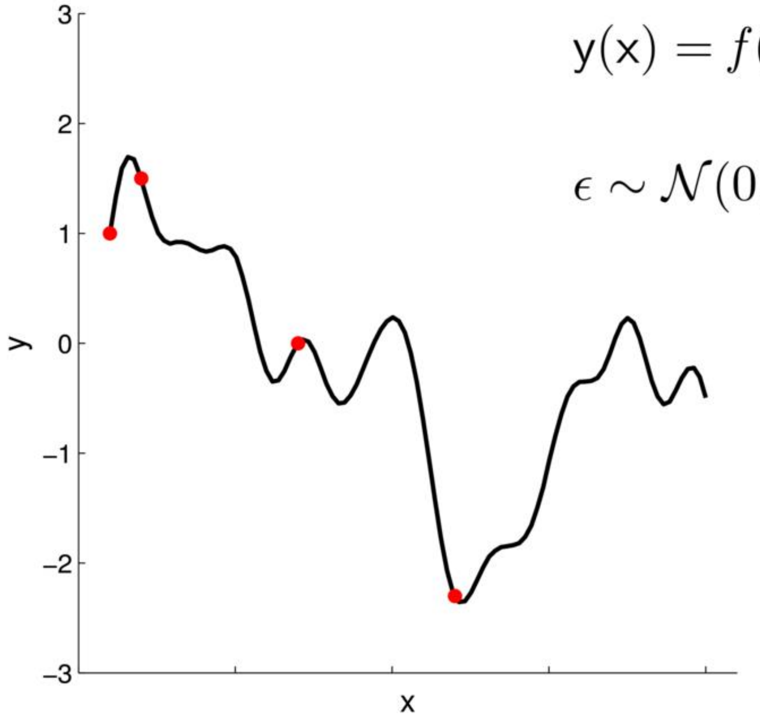
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

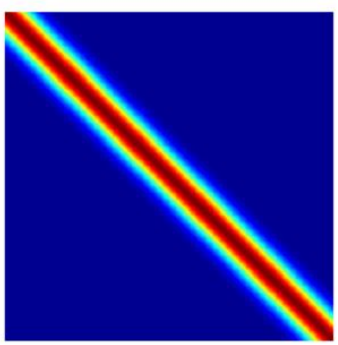
What effect do the Hyper-parameters Have?

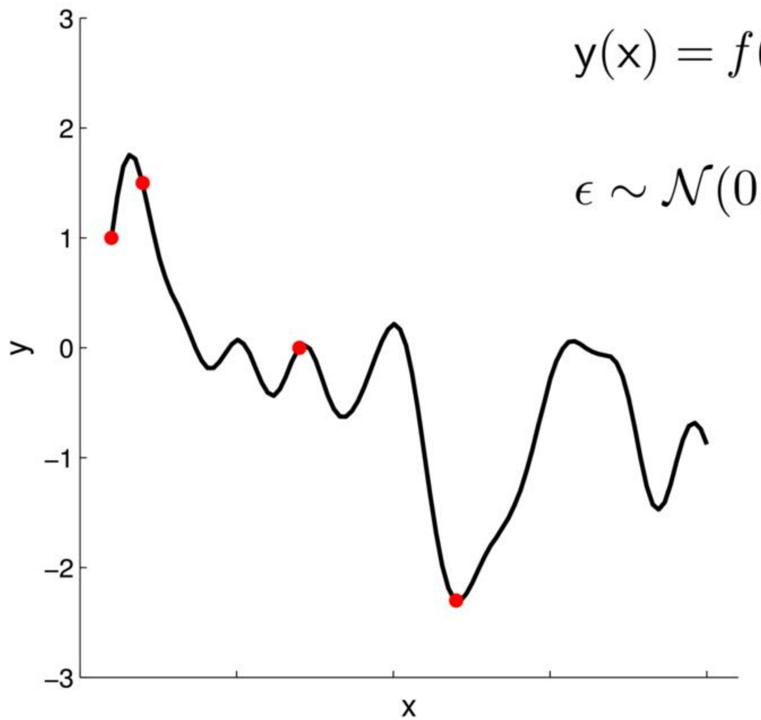
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

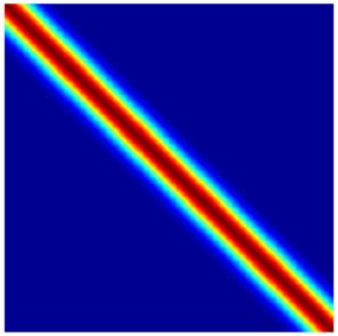
What effect do the Hyper-parameters Have?

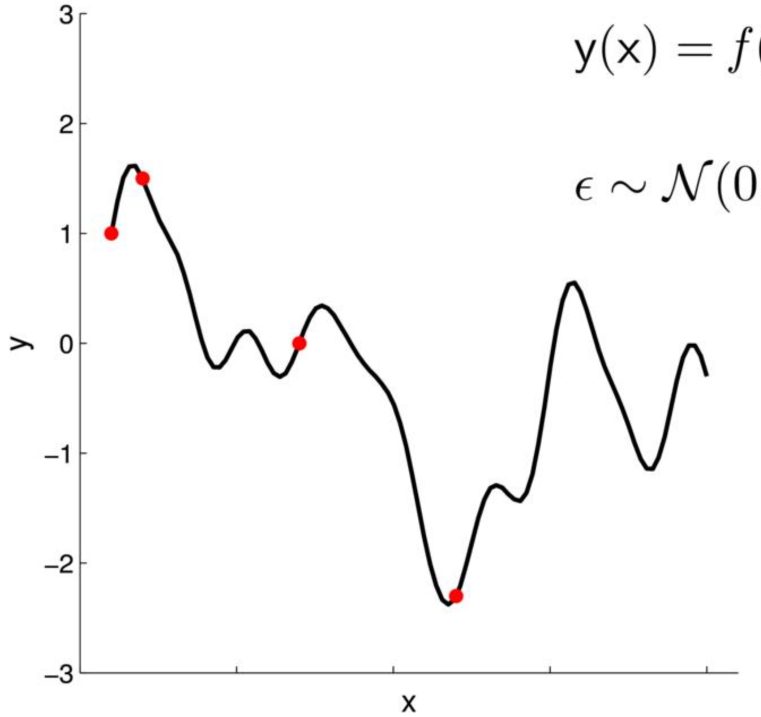
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

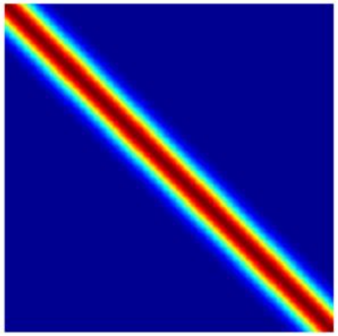
What effect do the Hyper-parameters Have?

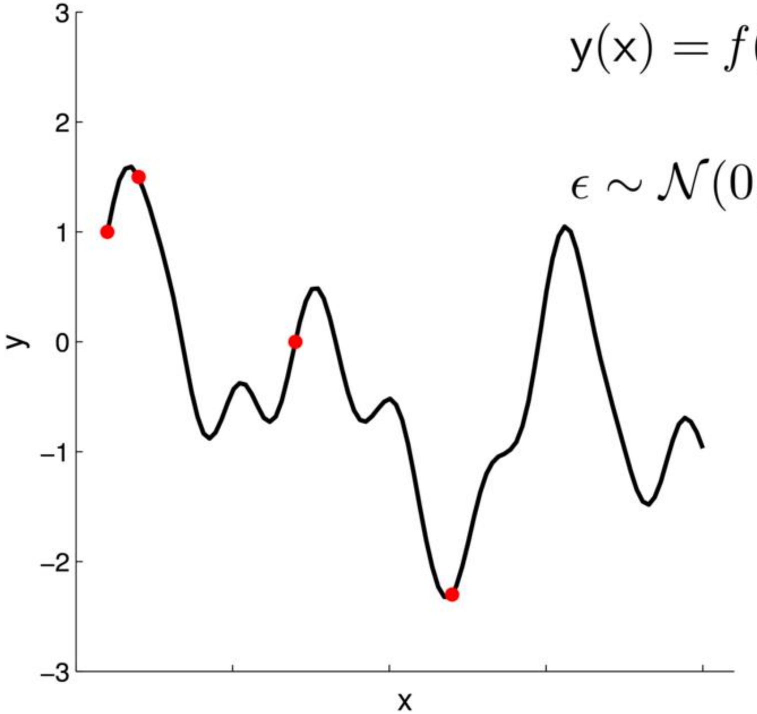
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

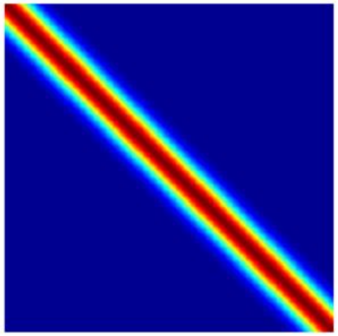
What effect do the Hyper-parameters Have?

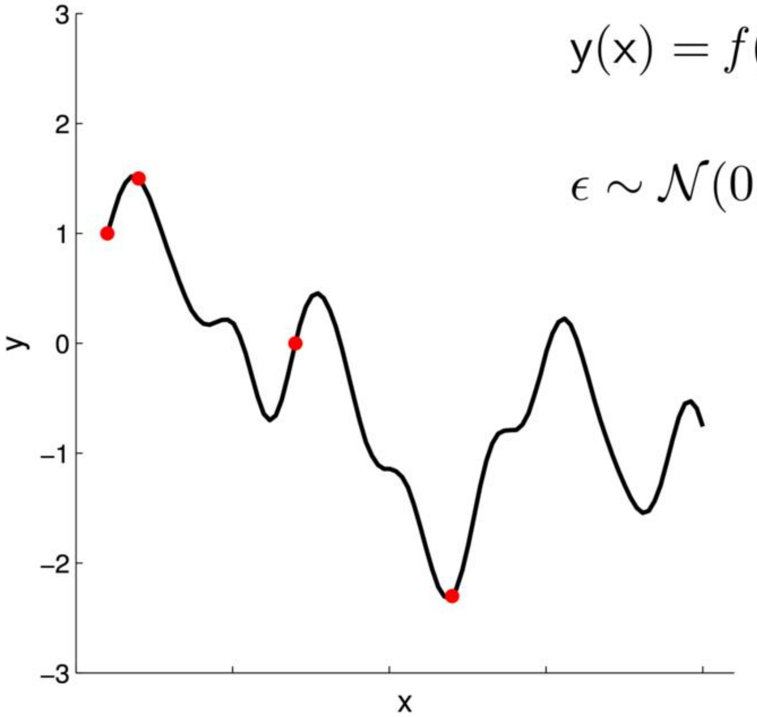
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

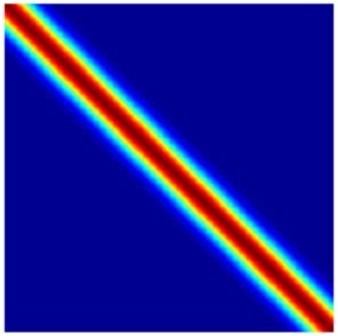
What effect do the Hyper-parameters Have?

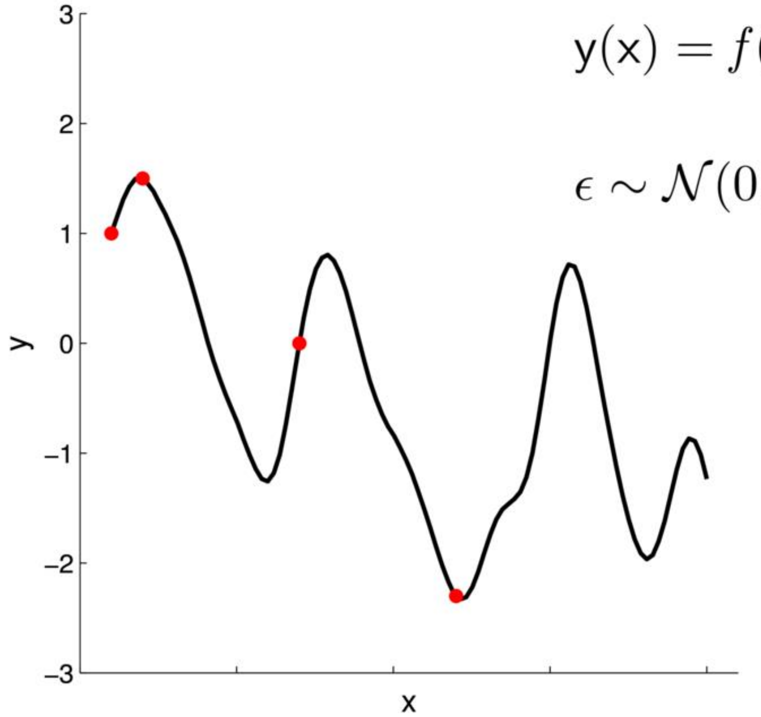
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

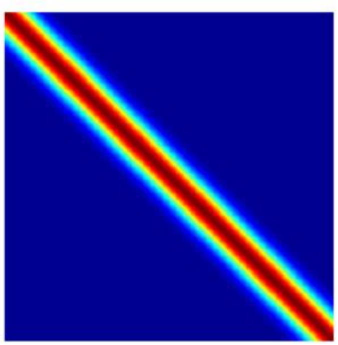
What effect do the Hyper-parameters Have?

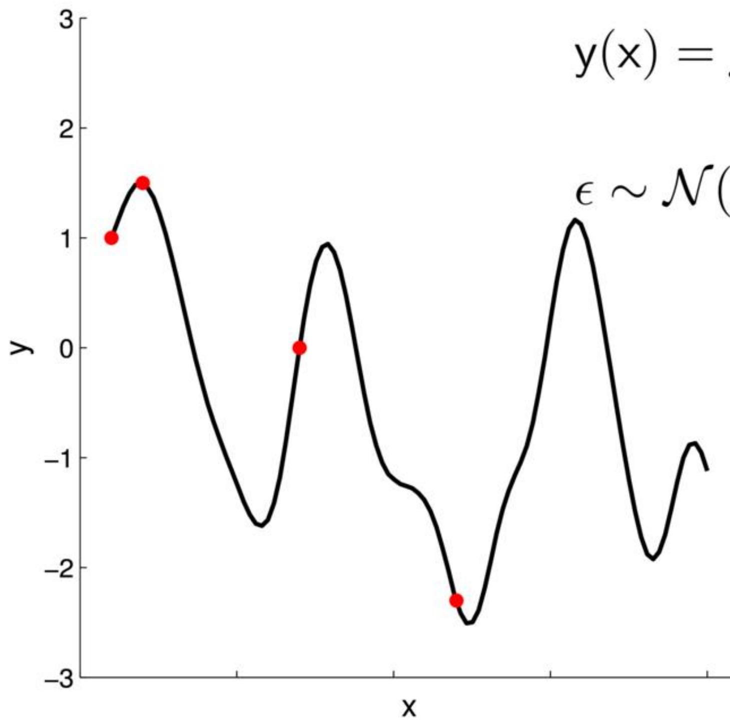
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

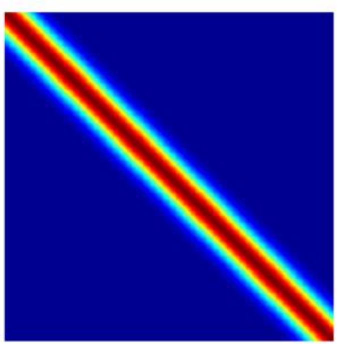
What effect do the Hyper-parameters Have?

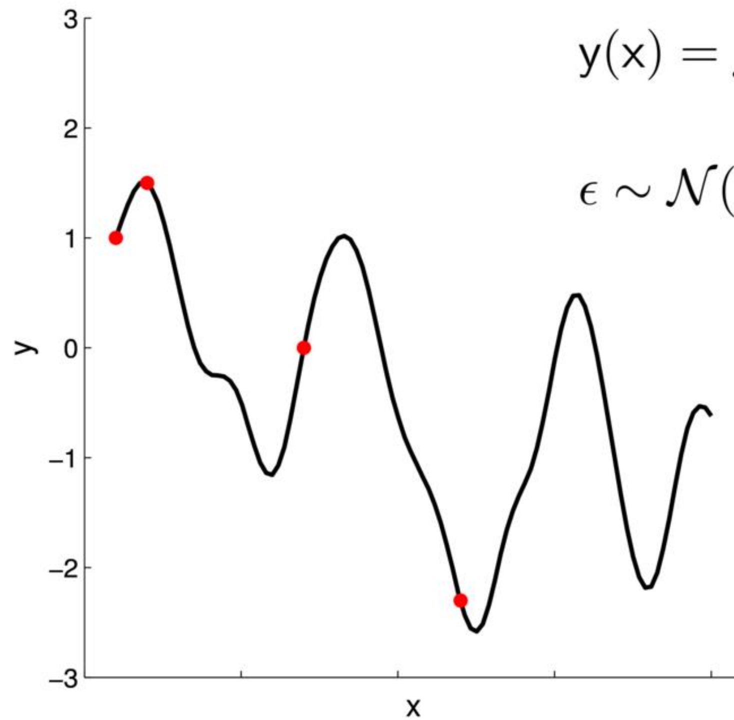
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

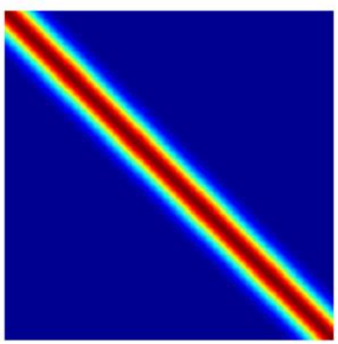
What effect do the Hyper-parameters Have?

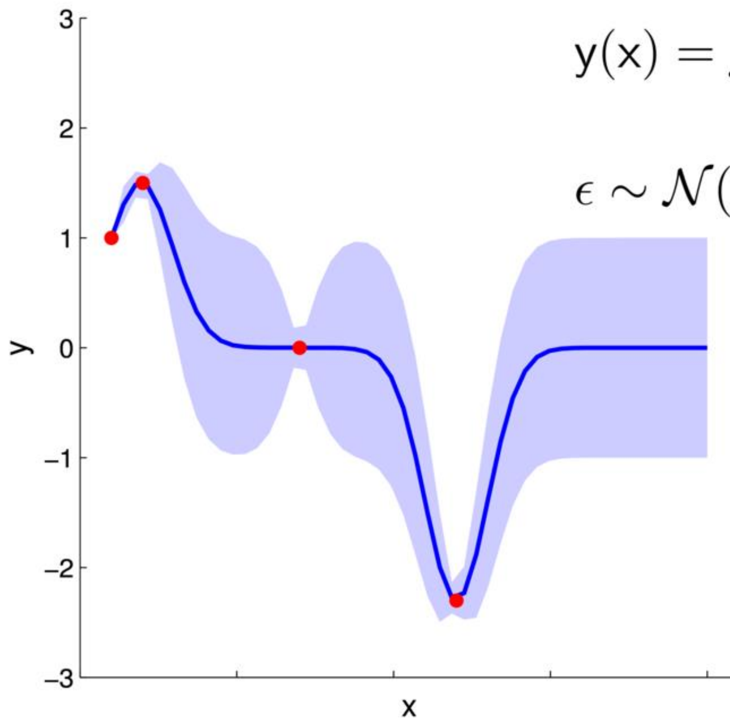
Non-parametric (∞-parametric)

$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$

$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$

$\Sigma =$ 



Parametric model

$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

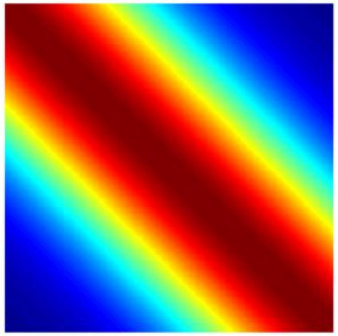
What effect do the Hyper-parameters Have?

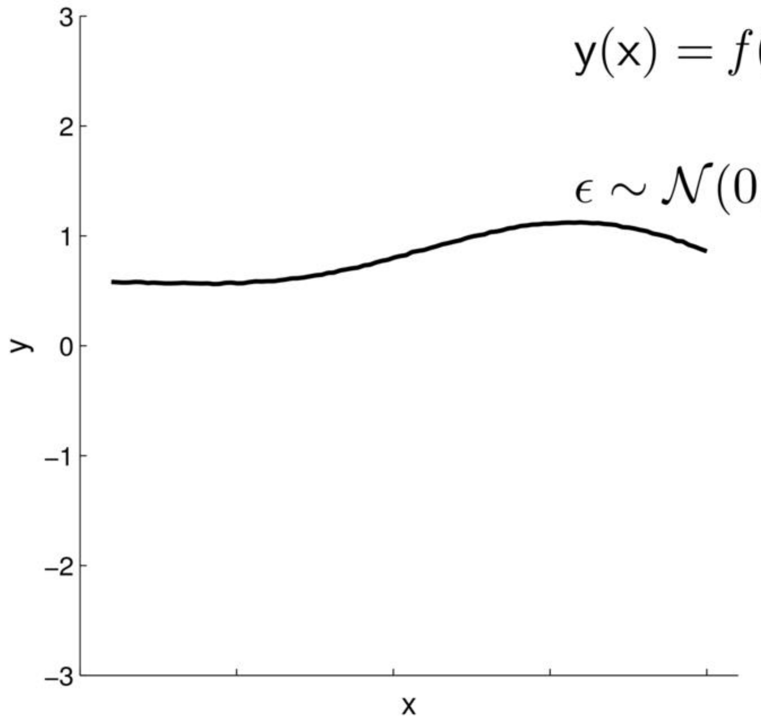
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

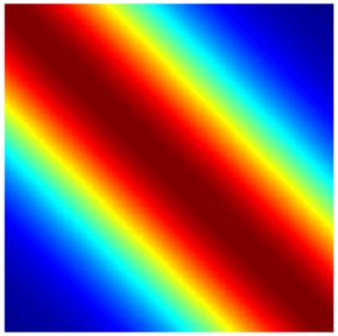
What effect do the Hyper-parameters Have?

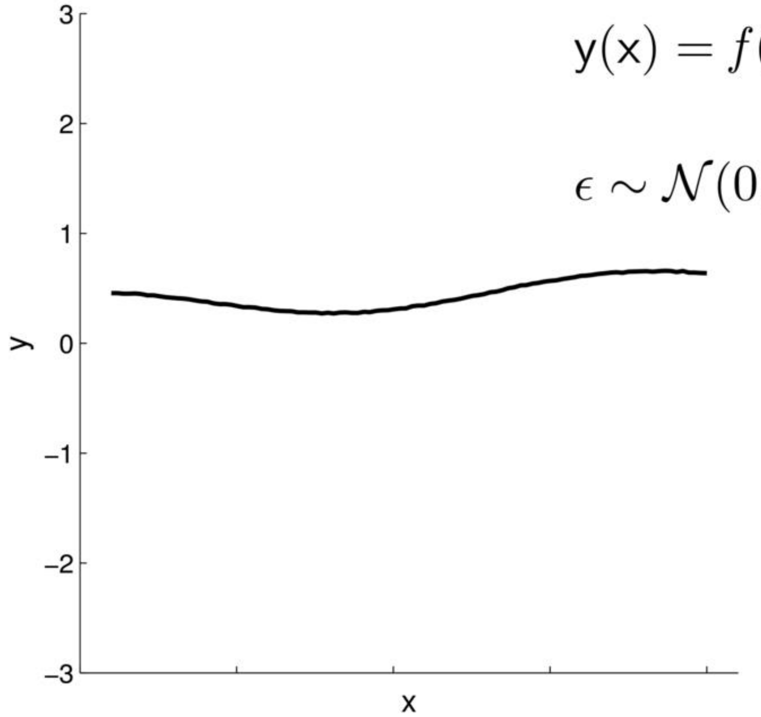
Non-parametric (∞-parametric)

$$p(y|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(x_1, x_2) = K(x_1, x_2) + I\sigma_y^2$$

$$K(x_1, x_2) = \sigma^2 e^{-\frac{1}{2l^2}(x_1 - x_2)^2}$$

$\Sigma =$ 



Parametric model

$$y(x) = f(x; \theta) + \sigma_y \epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the Hyper-parameters Have?

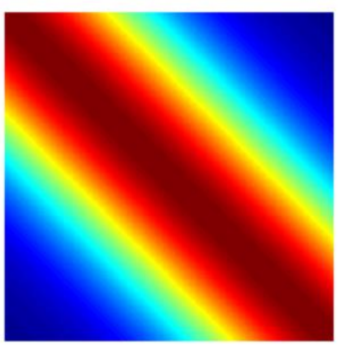
Non-parametric (∞-parametric)

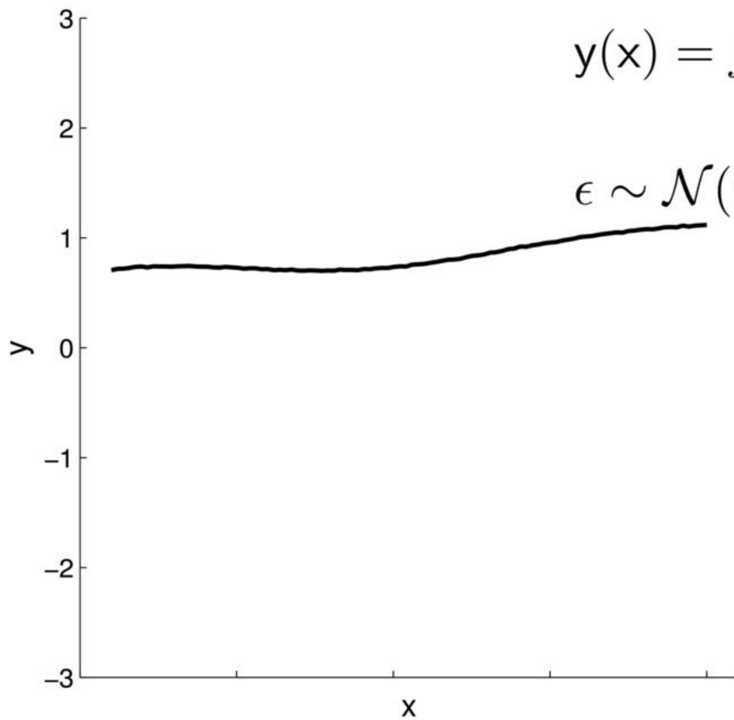
$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$





Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the Hyper-parameters Have?

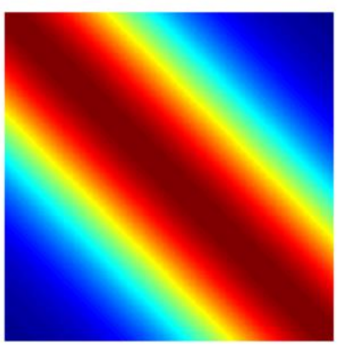
Non-parametric (∞-parametric)

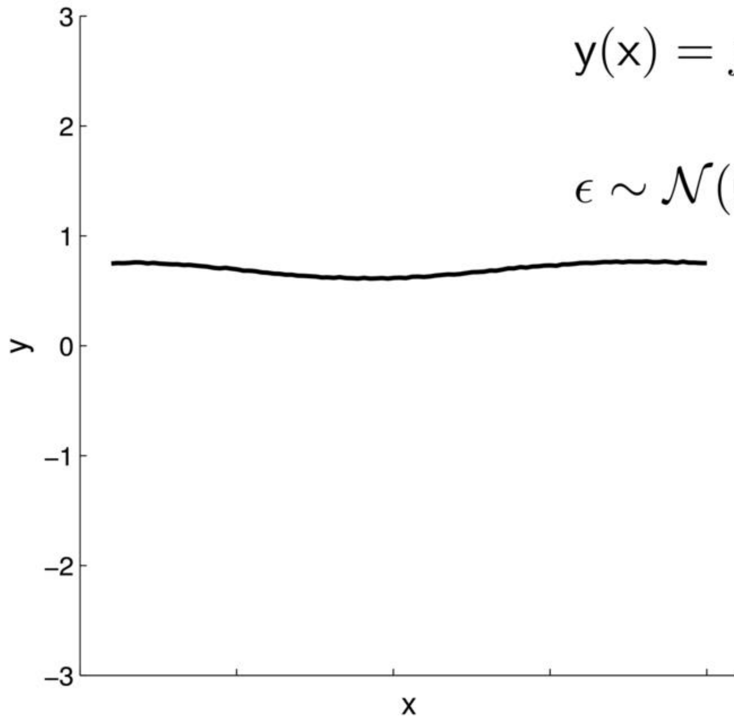
$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$





Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the Hyper-parameters Have?

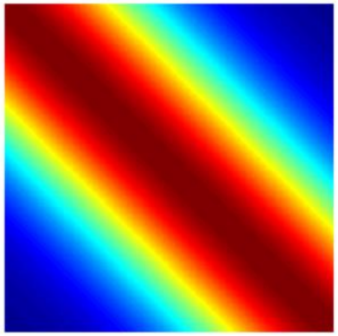
Non-parametric (∞-parametric)

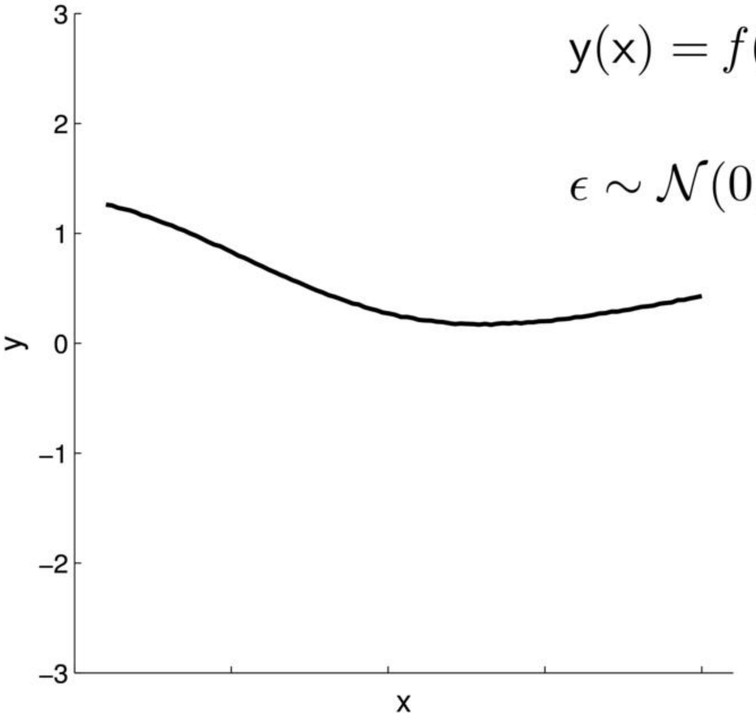
$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$

$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$

$\Sigma =$





Parametric model

$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

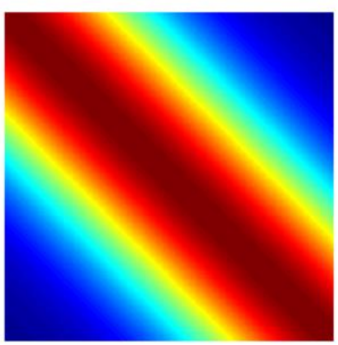
What effect do the Hyper-parameters Have?

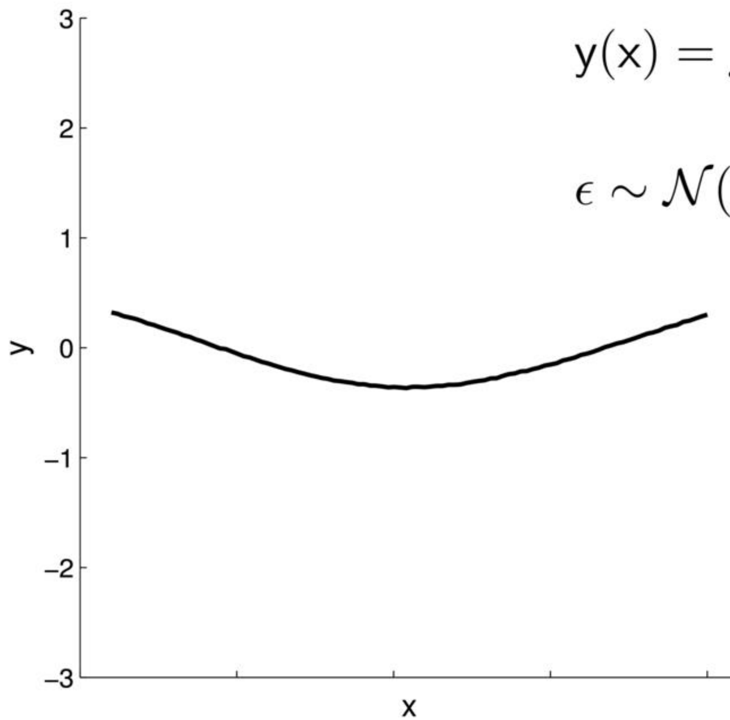
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

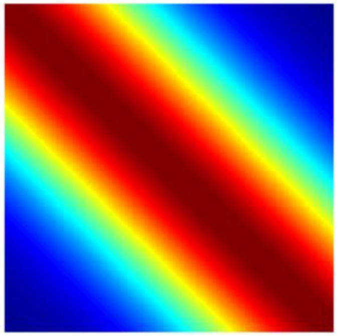
What effect do the Hyper-parameters Have?

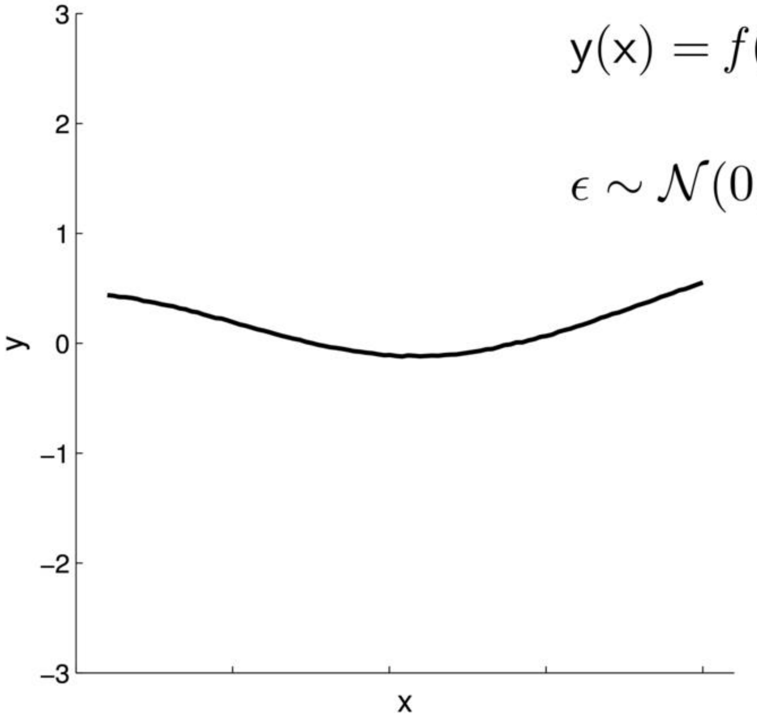
Non-parametric (∞-parametric)

$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$

$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$

$\Sigma =$ 



Parametric model

$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

What effect do the Hyper-parameters Have?

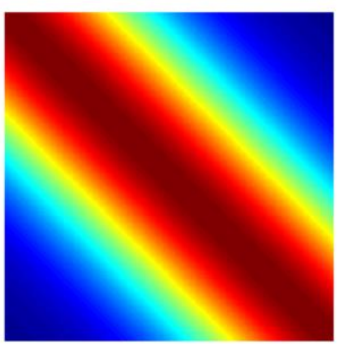
Non-parametric (∞-parametric)

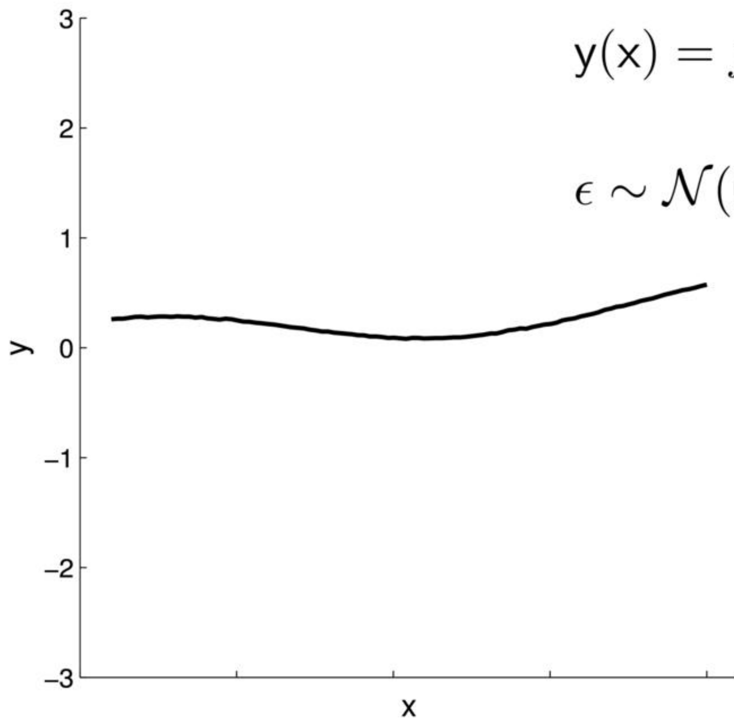
$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$





Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

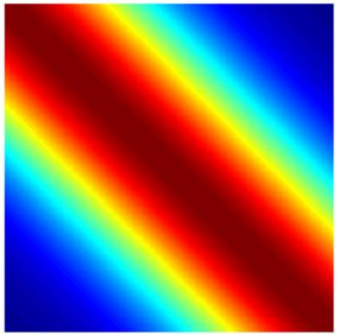
What effect do the Hyper-parameters Have?

Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

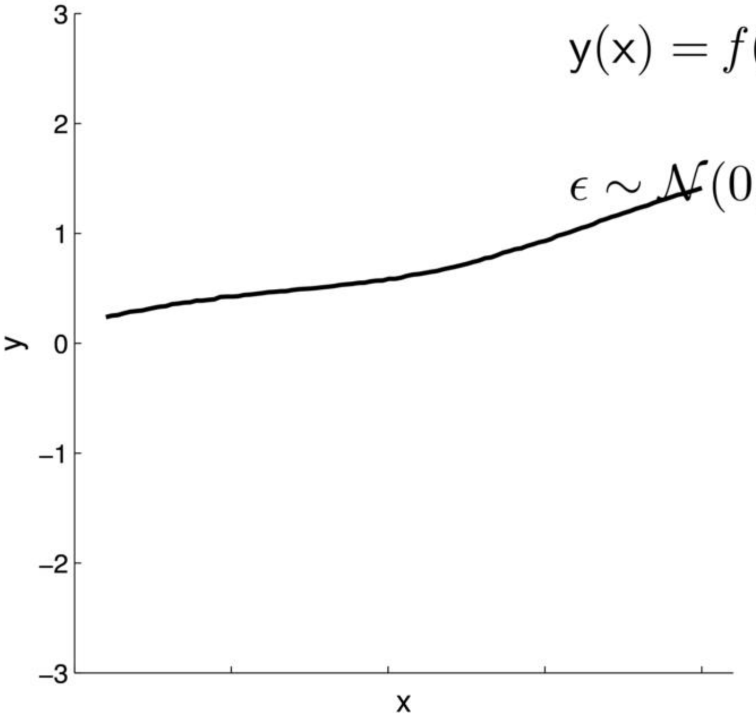
$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$



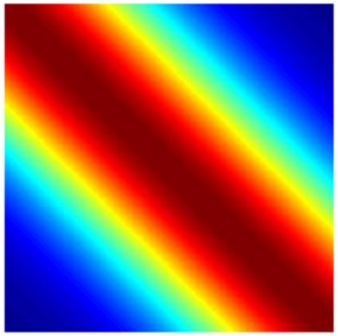
What effect do the Hyper-parameters Have?

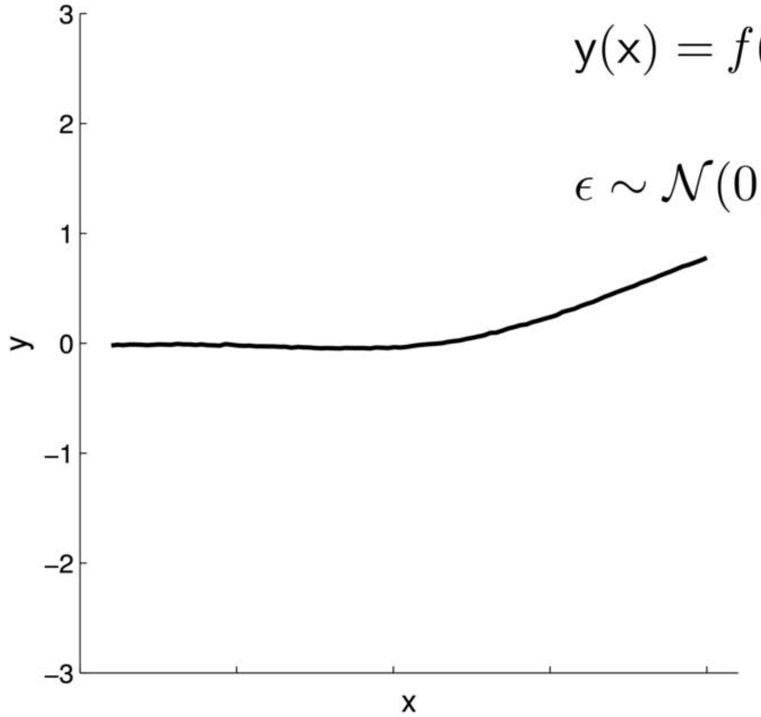
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

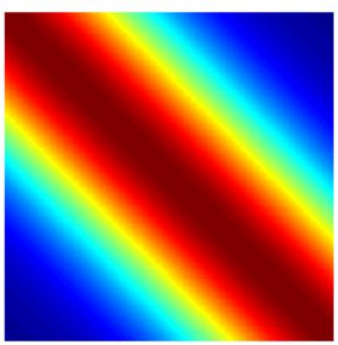
What effect do the Hyper-parameters Have?

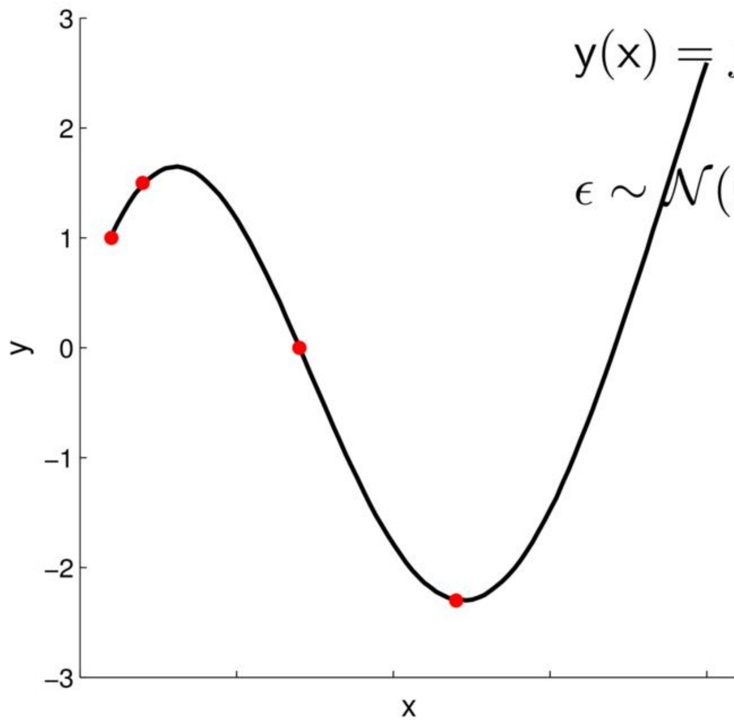
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

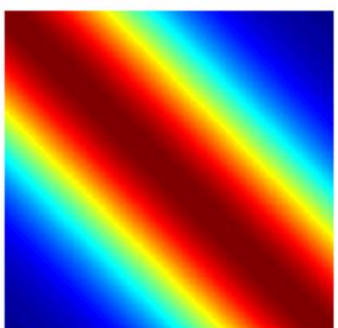
What effect do the Hyper-parameters Have?

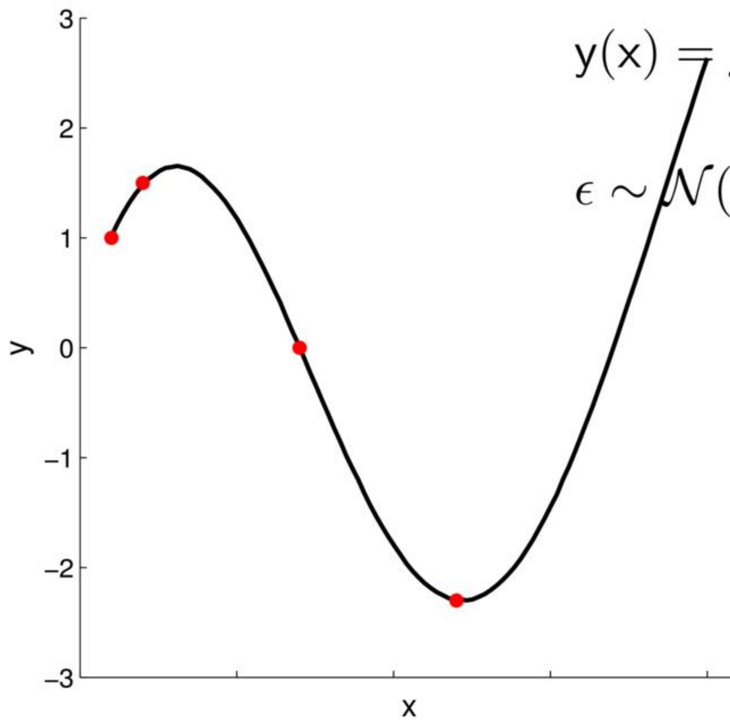
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

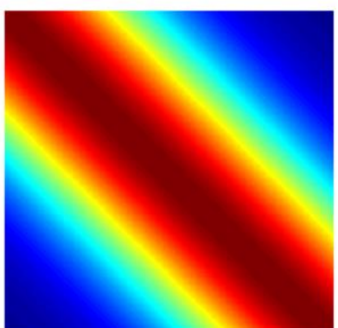
What effect do the Hyper-parameters Have?

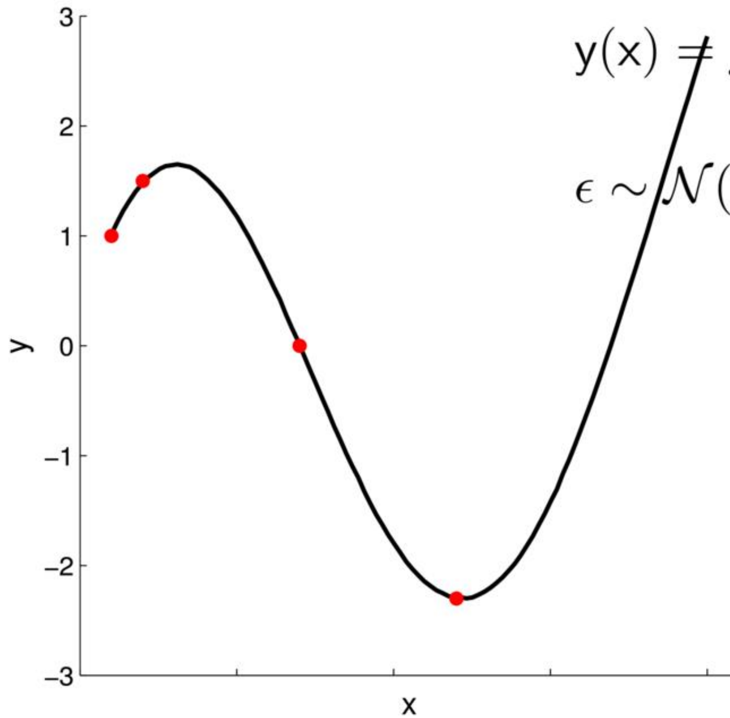
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

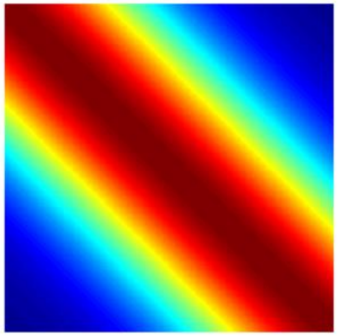
What effect do the Hyper-parameters Have?

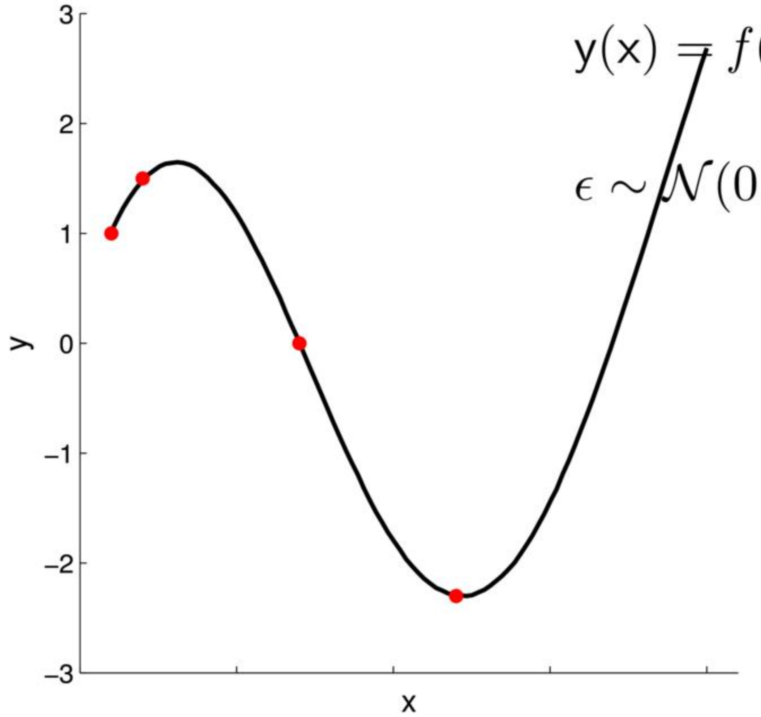
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

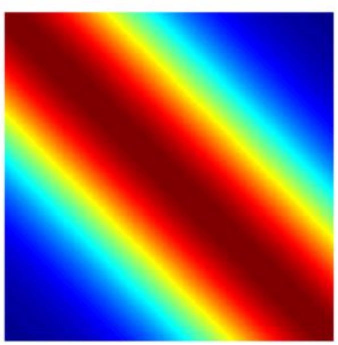
What effect do the Hyper-parameters Have?

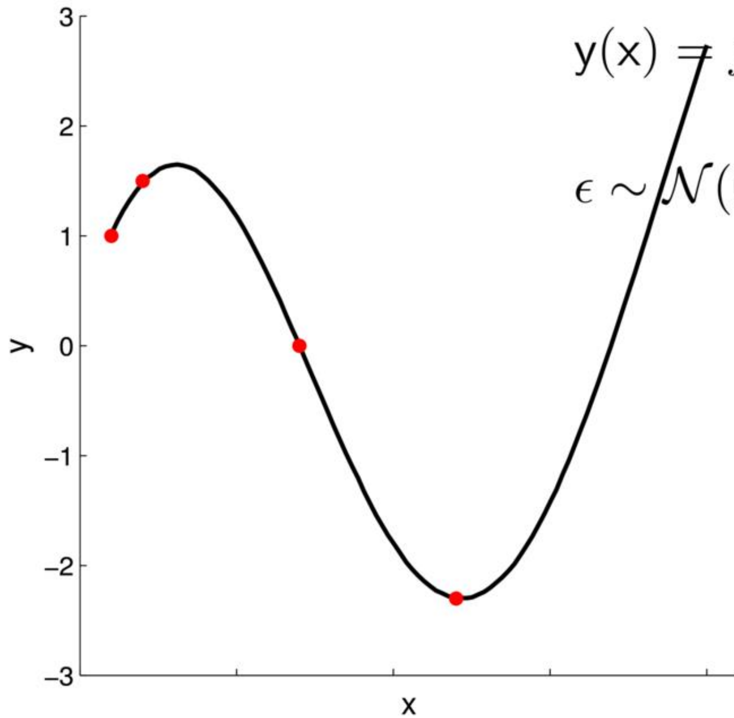
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the Hyper-parameters Have?

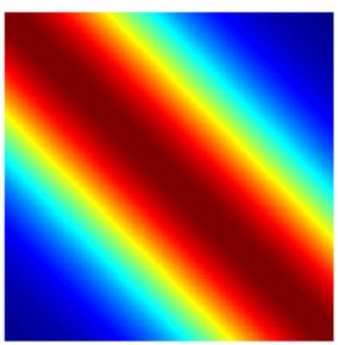
Non-parametric (∞-parametric)

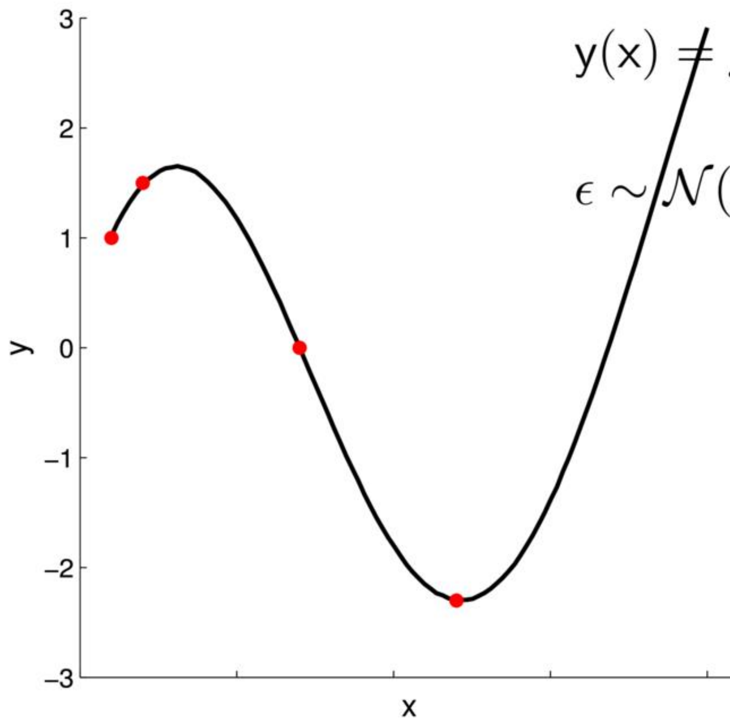
$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$

$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$

$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$

$\Sigma =$





Parametric model

$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

What effect do the Hyper-parameters Have?

Non-parametric (∞-parametric)

Parametric model

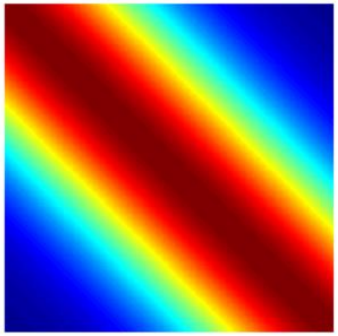
$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

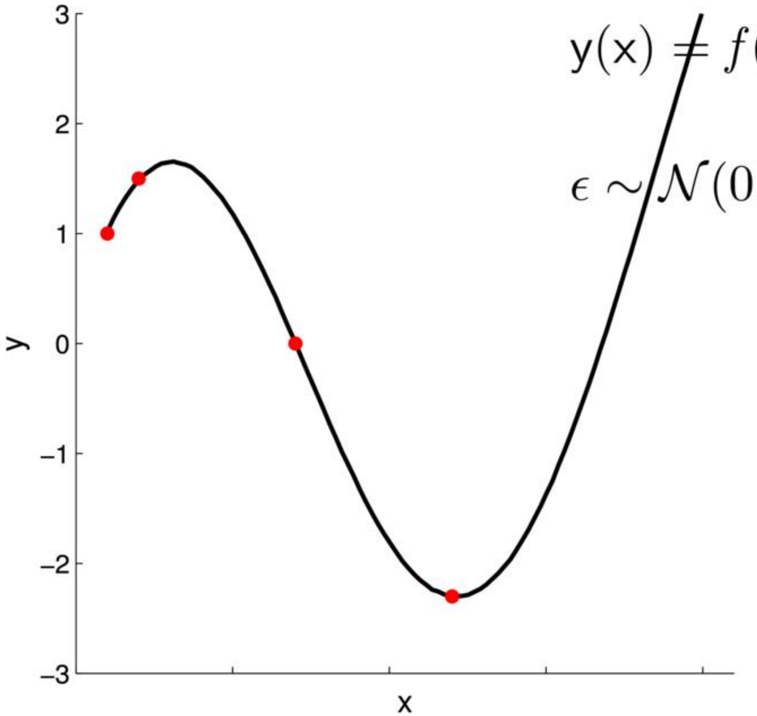
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}} \epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

$\Sigma =$ 



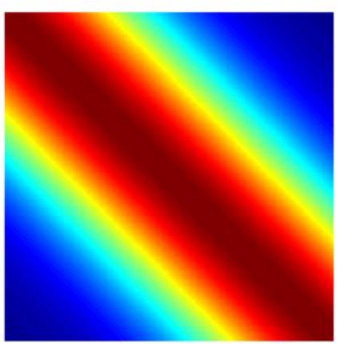
What effect do the Hyper-parameters Have?

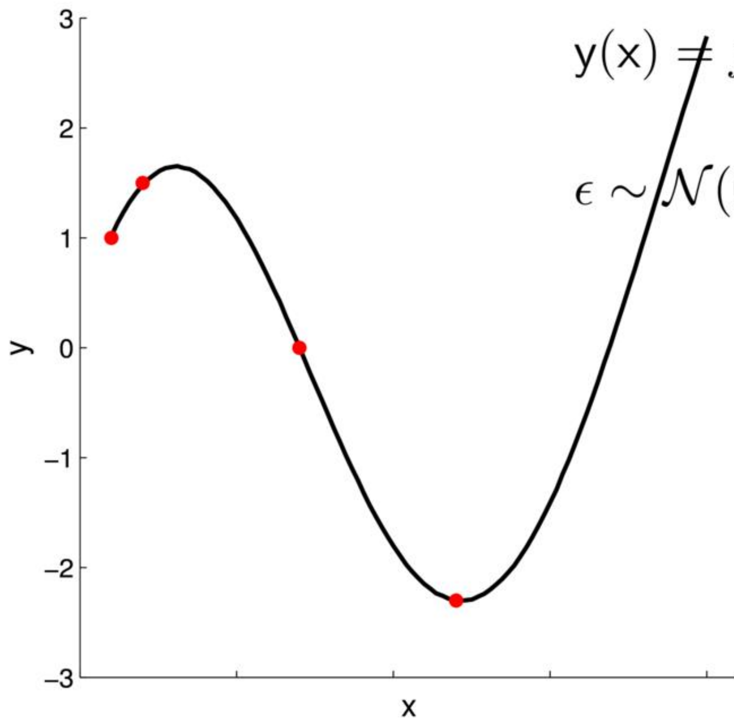
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

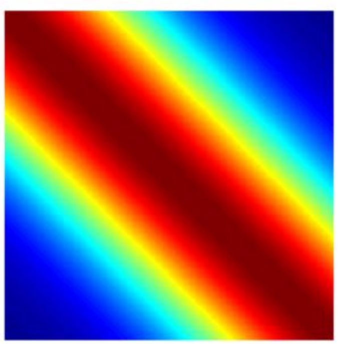
What effect do the Hyper-parameters Have?

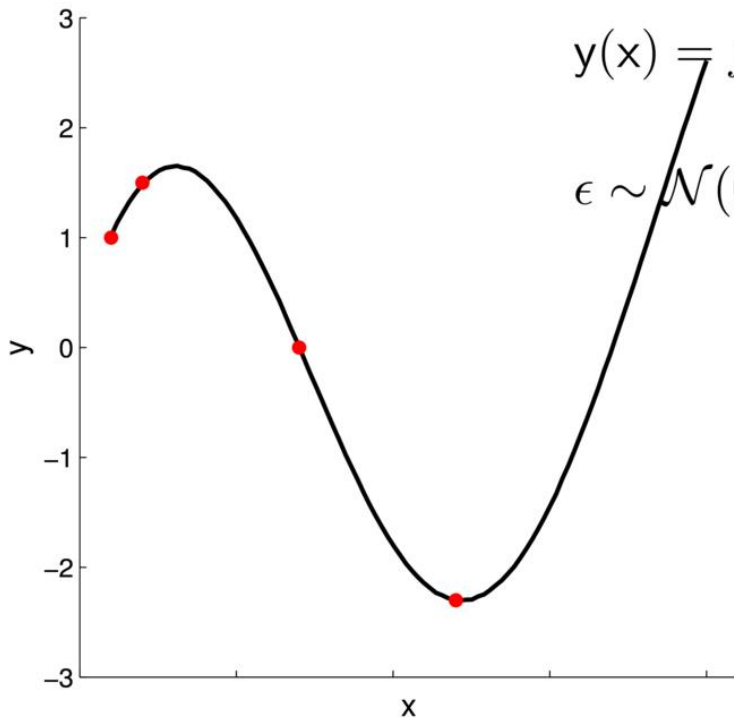
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the Hyper-parameters Have?

Non-parametric (∞-parametric)

Parametric model

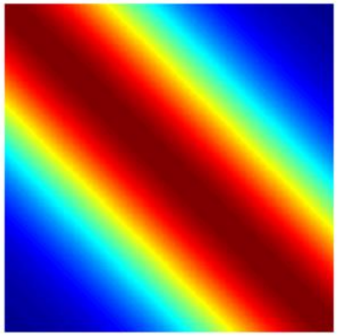
$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$

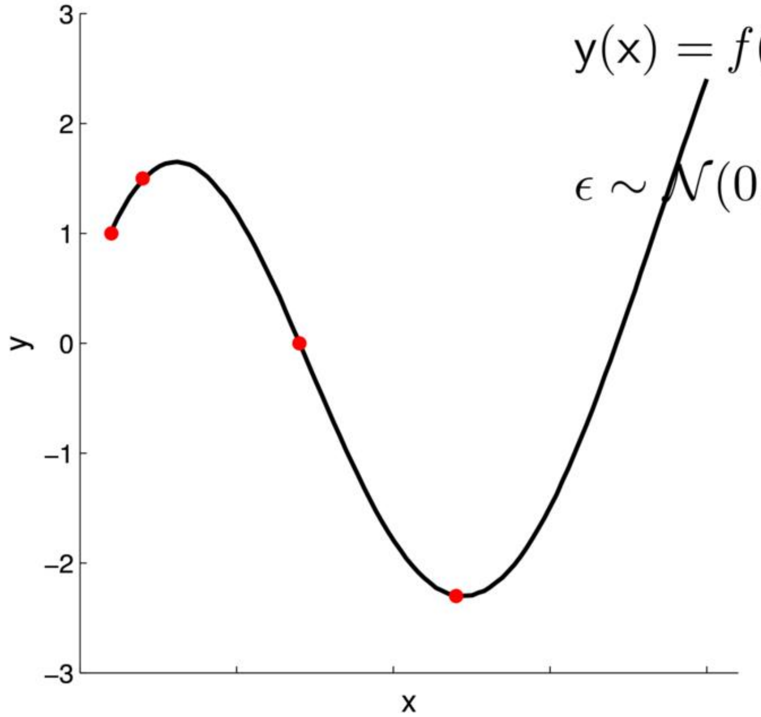
$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$

$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$

$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$

$\epsilon \sim \mathcal{N}(0, 1)$

$\Sigma =$ 



What effect do the Hyper-parameters Have?

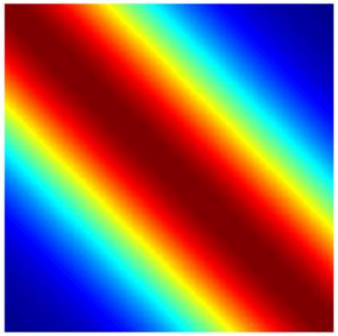
Non-parametric (∞-parametric)

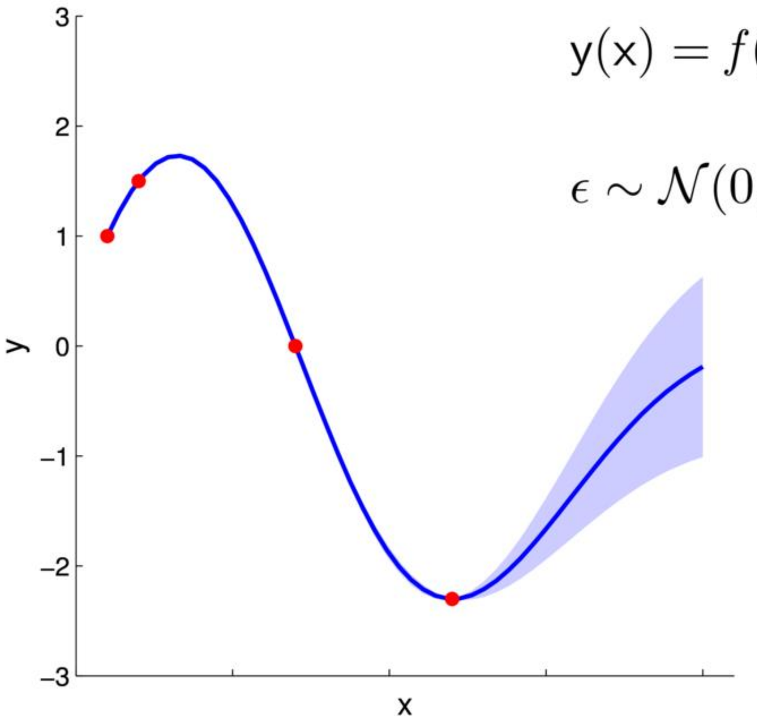
$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$





Parametric model

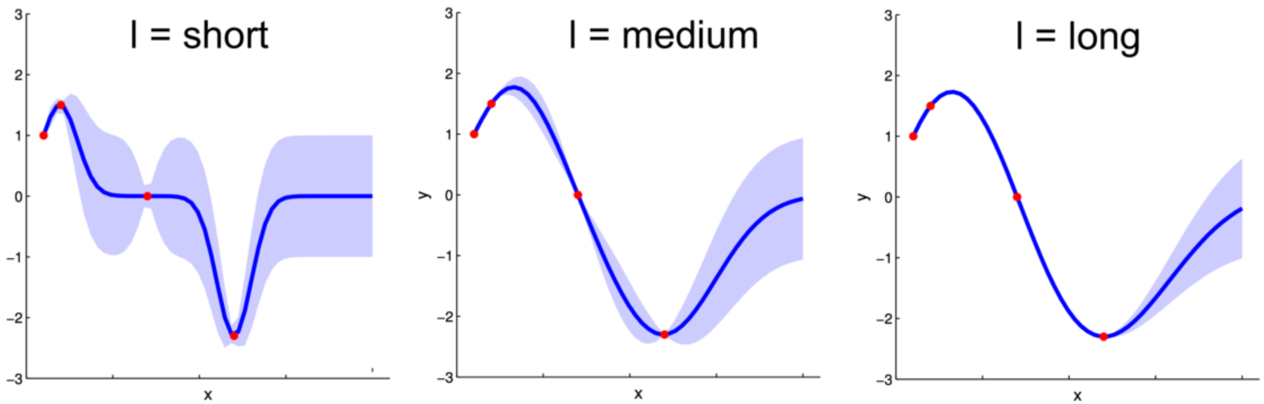
$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the Hyper-parameters Have?

$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

- Hyper-parameters have a strong effect
  - l controls the horizontal length-scale
  - σ^2 controls the vertical scale of the data
- $\Longrightarrow$ need automatic learning of hyper-parameters from data



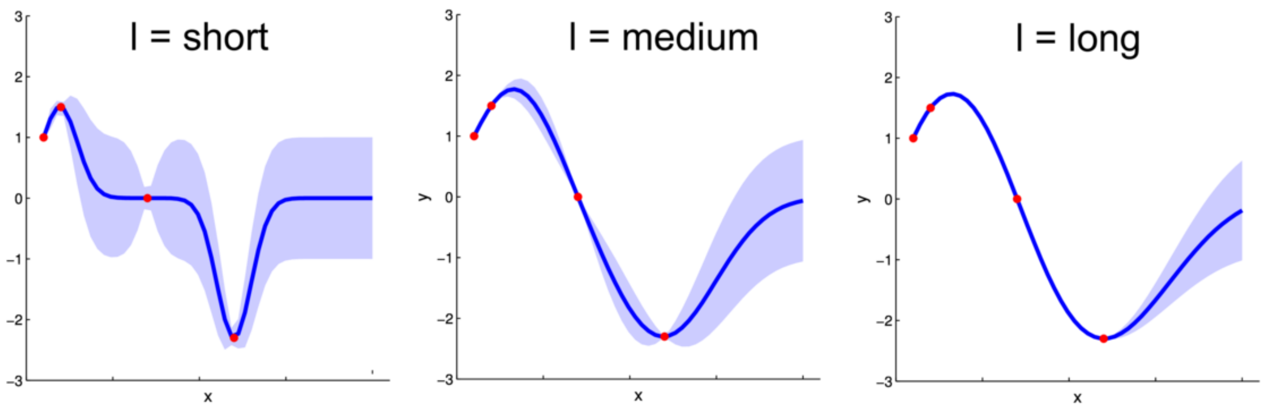
What effect do the Hyper-parameters Have?

$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

- Hyper-parameters have a strong effect
  - l controls the horizontal length-scale
  - σ^2 controls the vertical scale of the data
- $\Longrightarrow$ need automatic learning of hyper-parameters from data e.g.

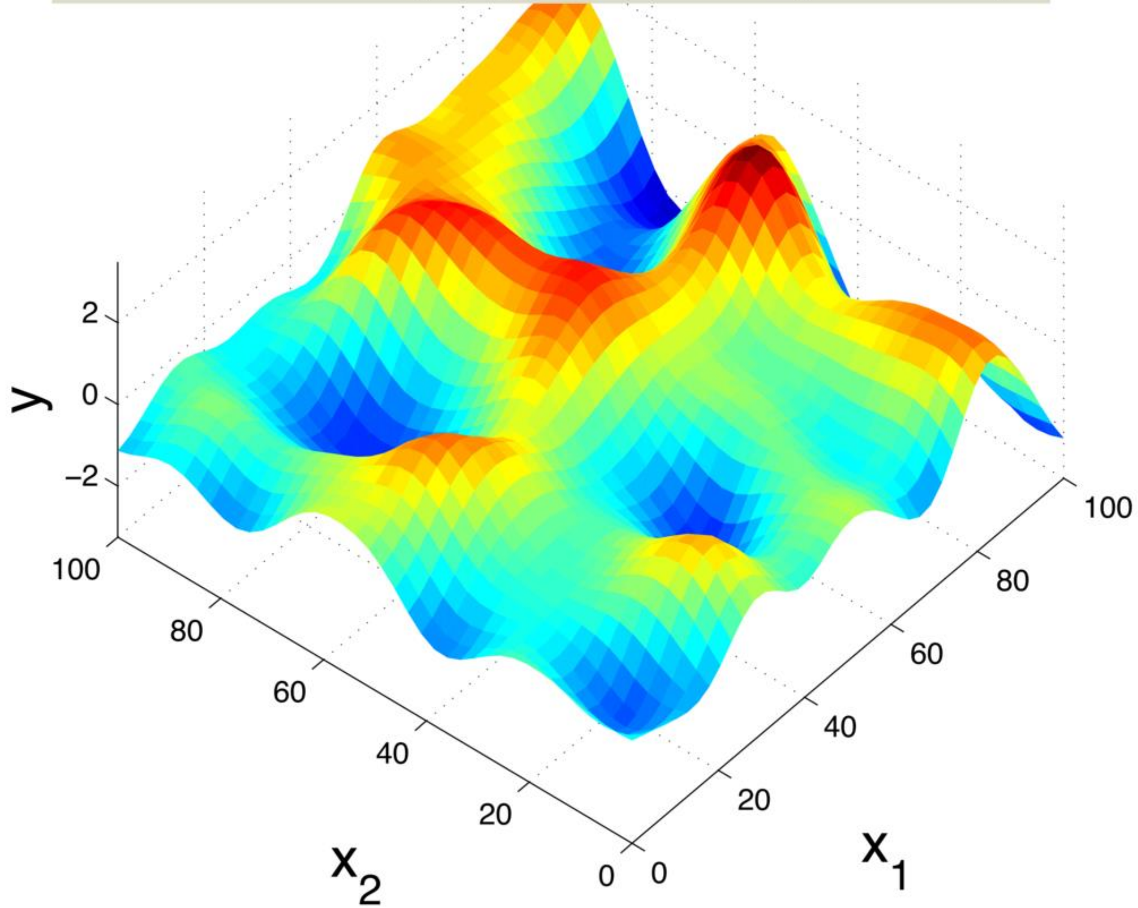
$$\arg\max_{l,\sigma^2} \log p(\mathbf{y} \mid \theta)$$

🡸 More on this later.

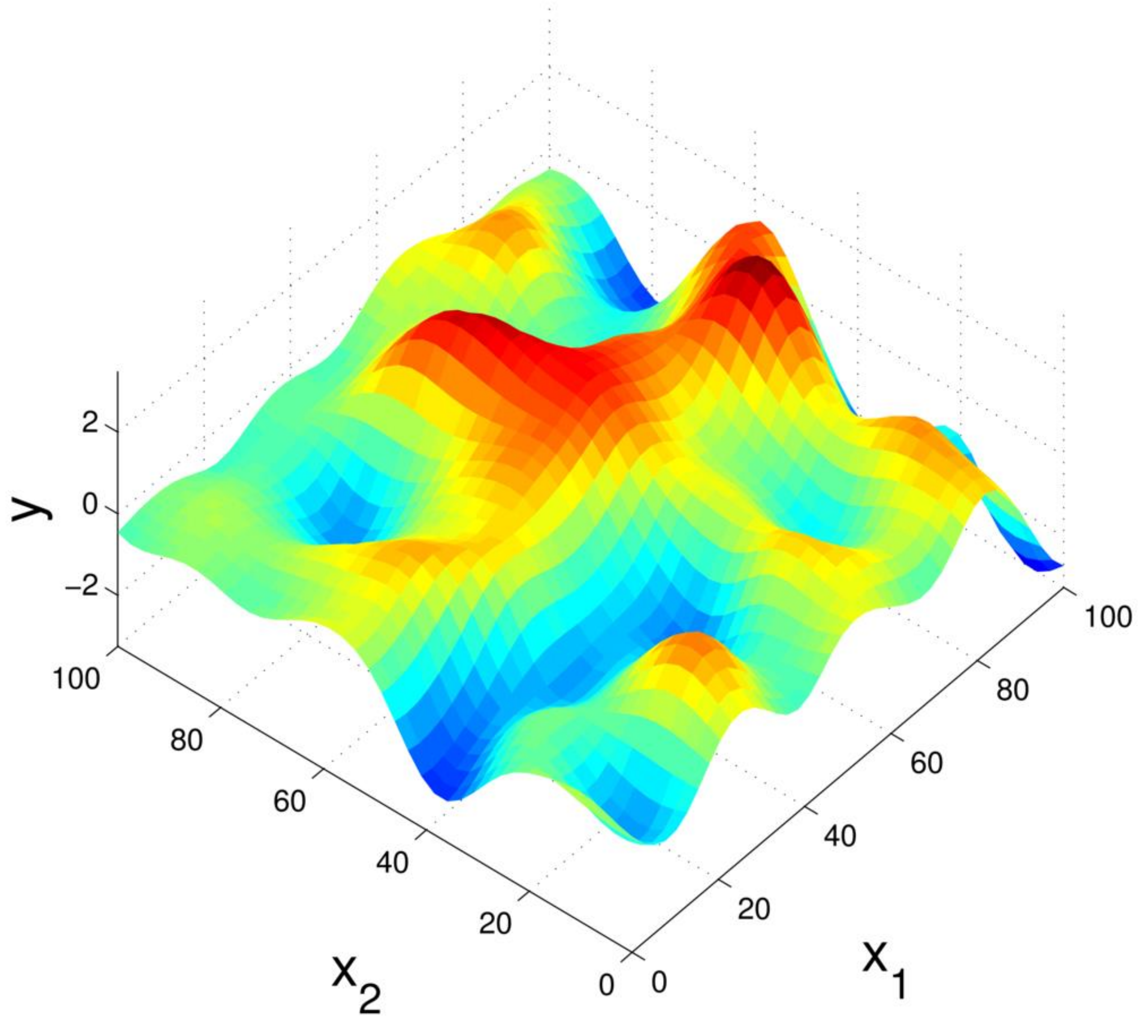


Higher Dimensional Input Spaces

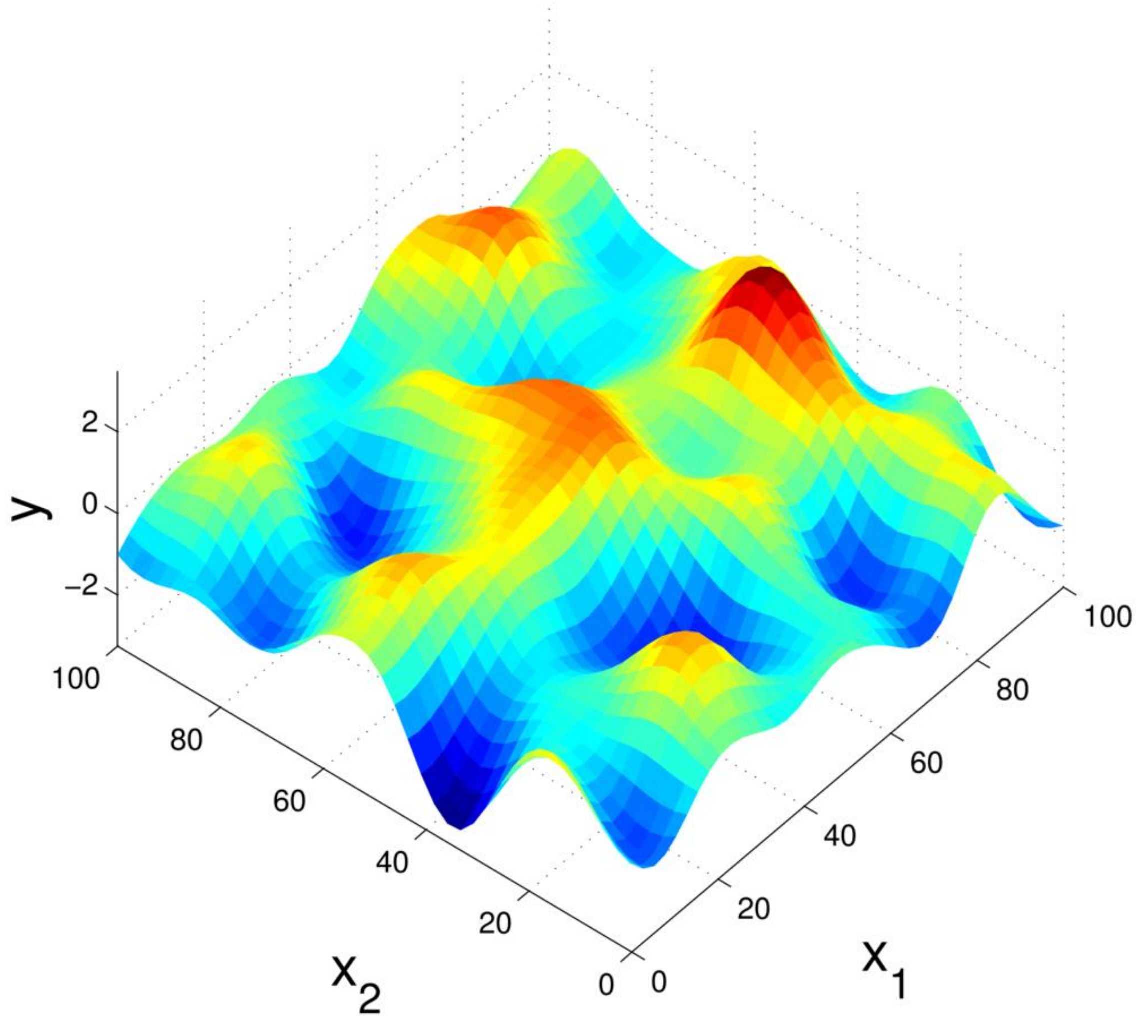
$$K(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{1}{2l_1^2}(\mathrm{x}_1 - \mathrm{x}_1')^2 - \frac{1}{2l_2^2}(\mathrm{x}_2 - \mathrm{x}_2')^2\right)$$



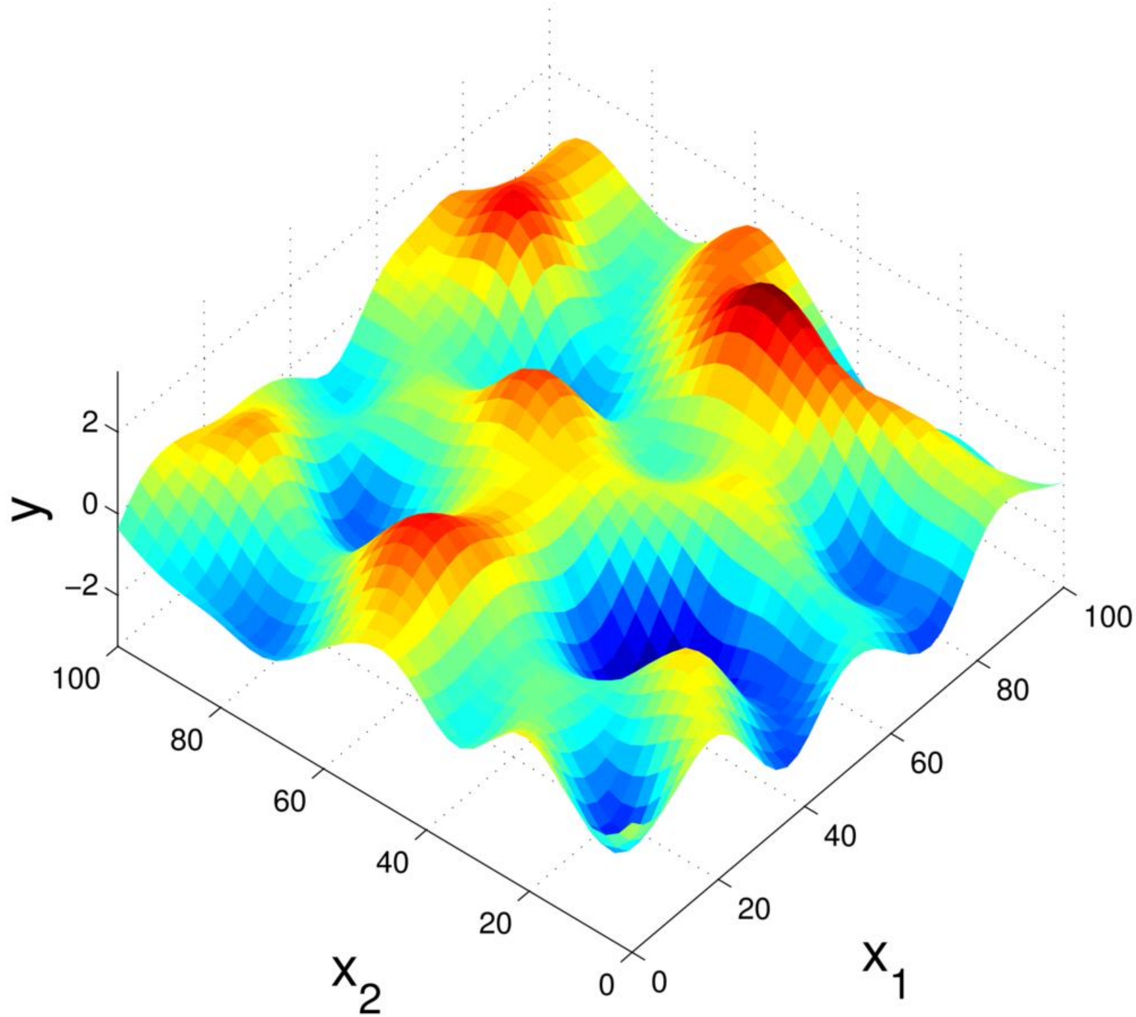
Higher Dimensional Input Spaces



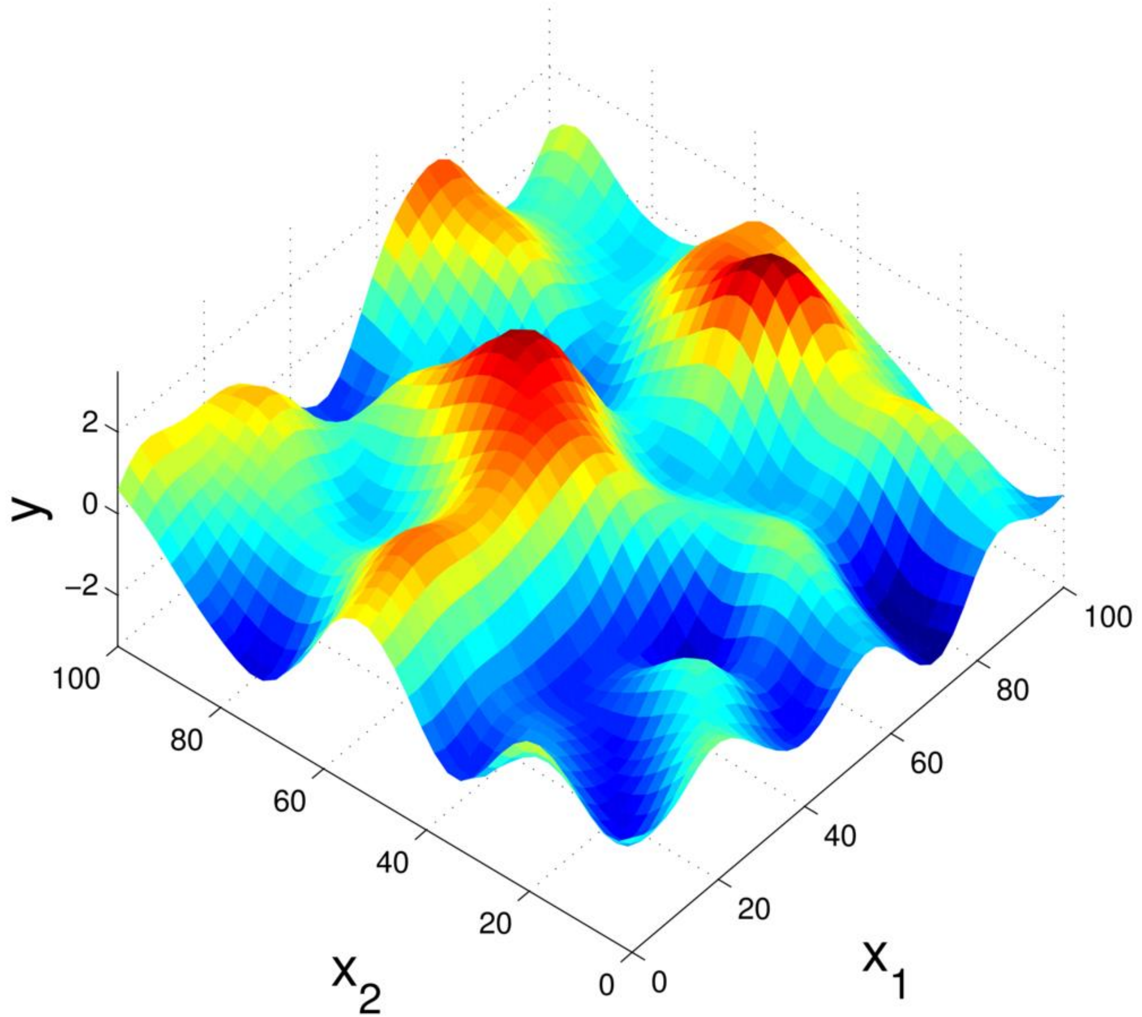
Higher Dimensional Input Spaces



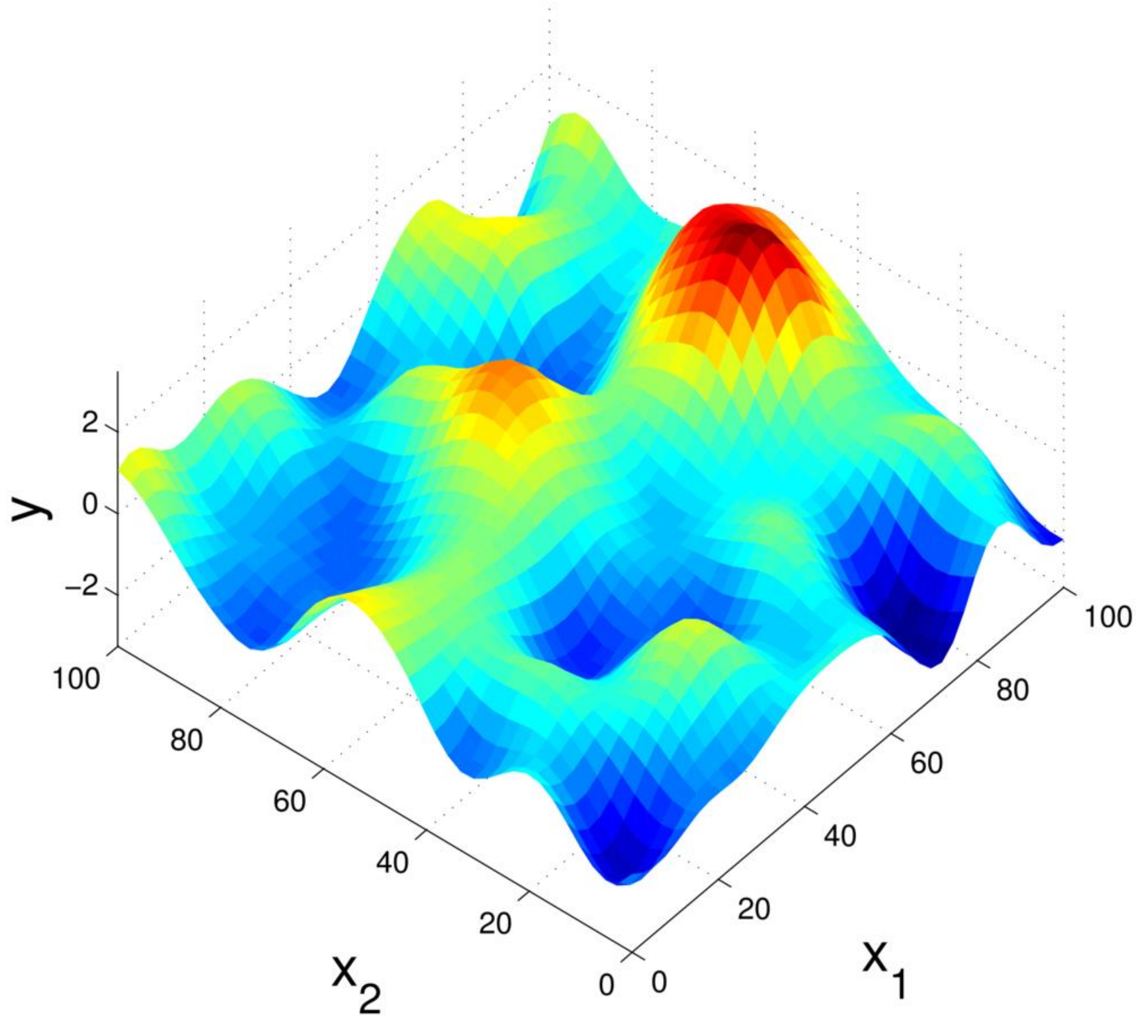
Higher Dimensional Input Spaces



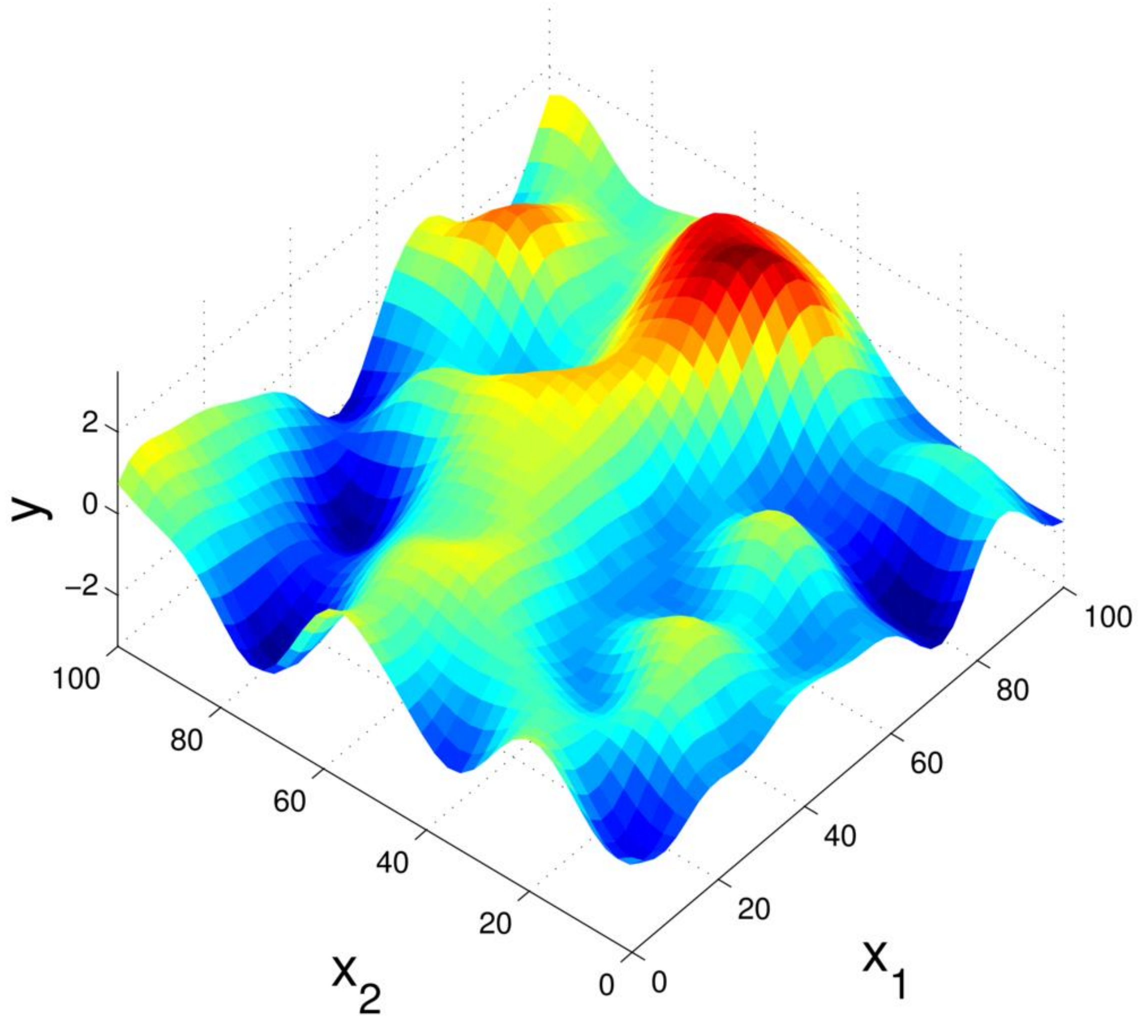
Higher Dimensional Input Spaces



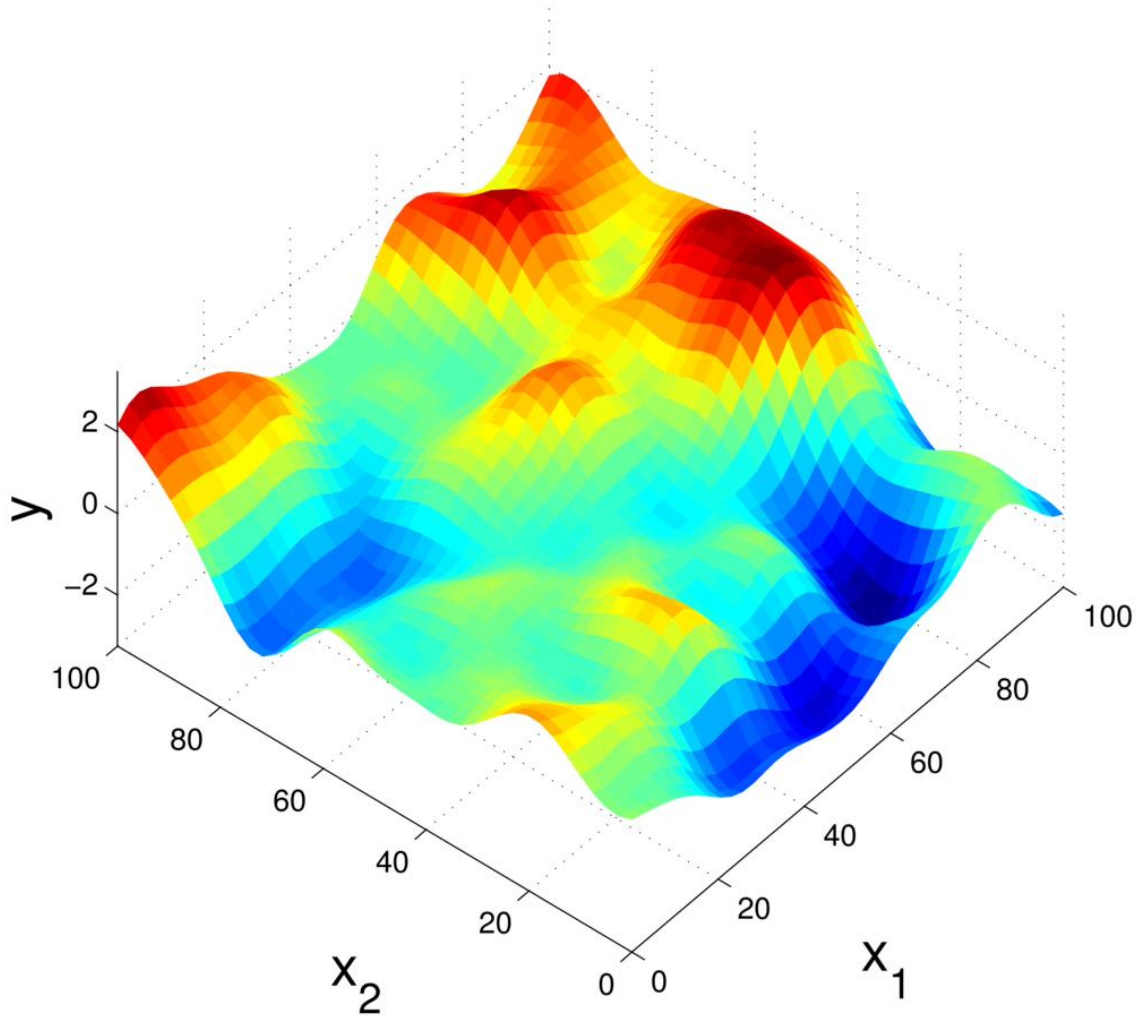
Higher Dimensional Input Spaces



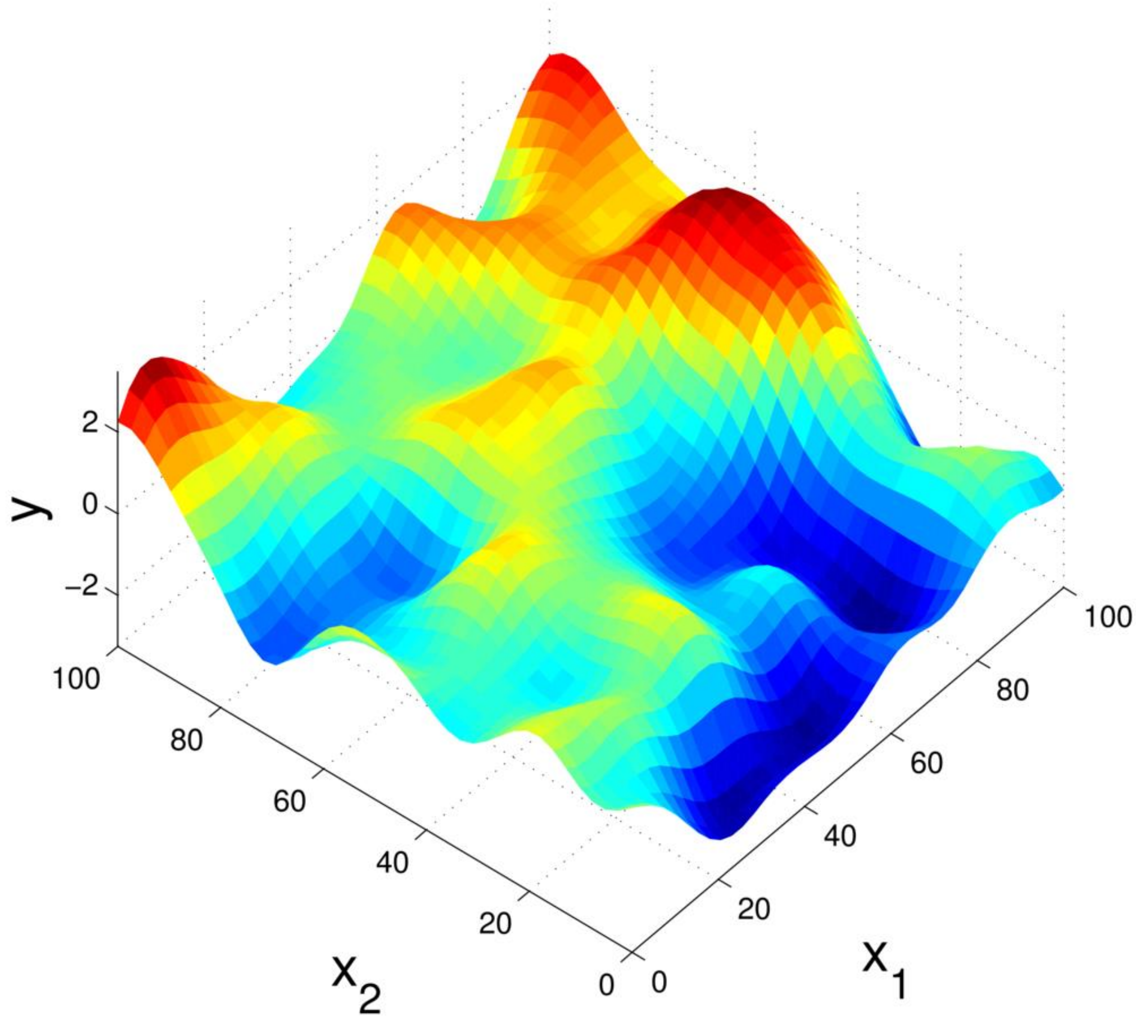
Higher Dimensional Input Spaces



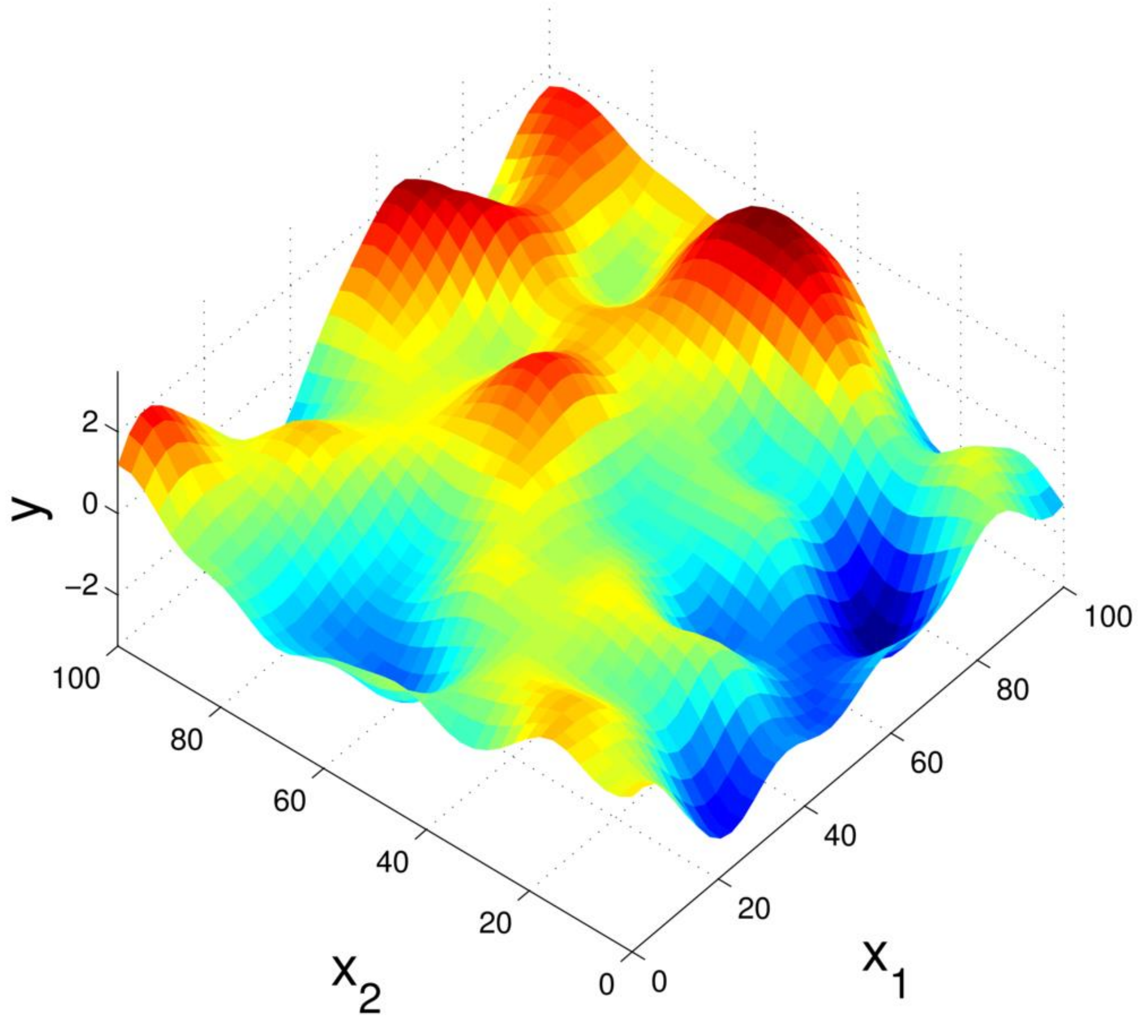
Higher Dimensional Input Spaces



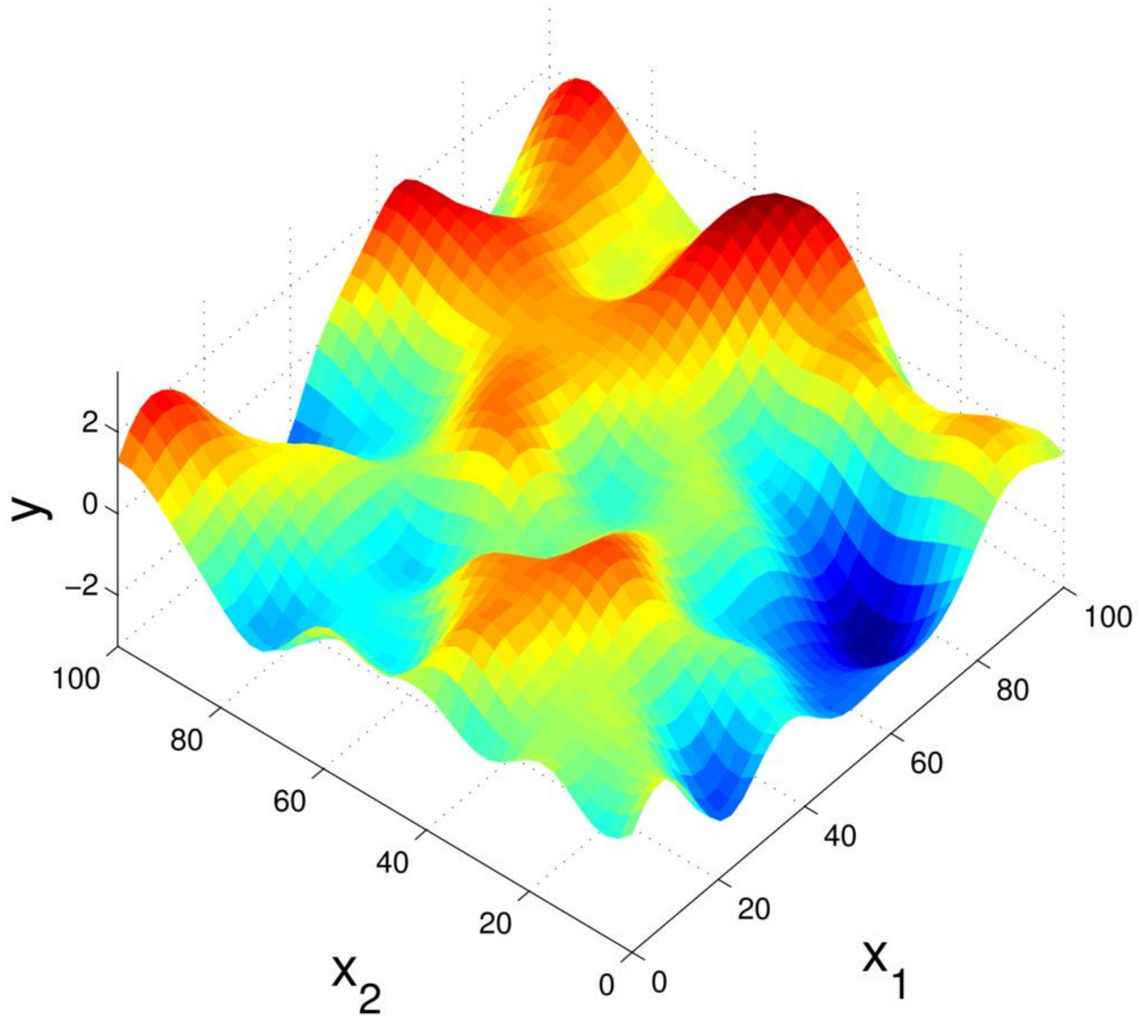
Higher Dimensional Input Spaces



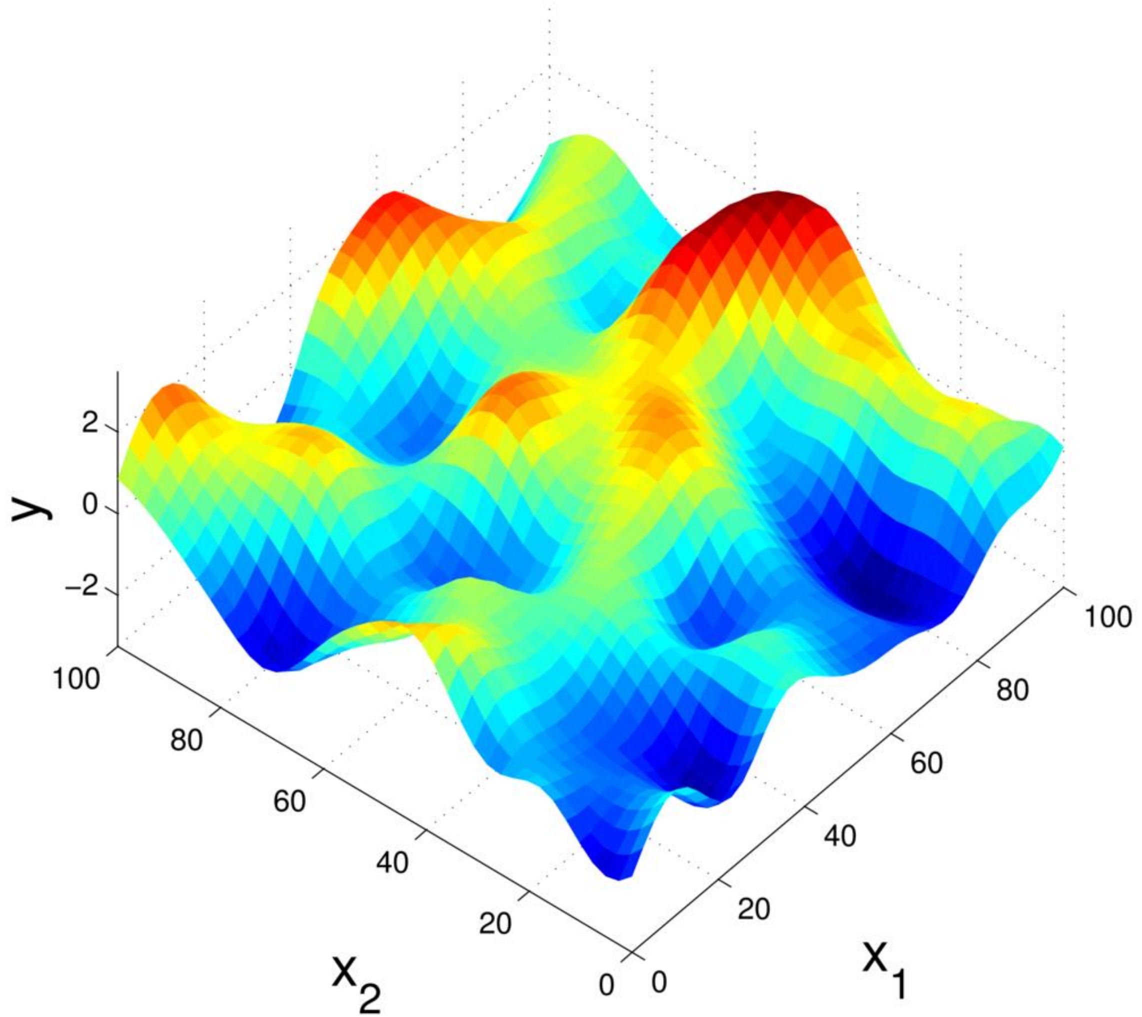
Higher Dimensional Input Spaces



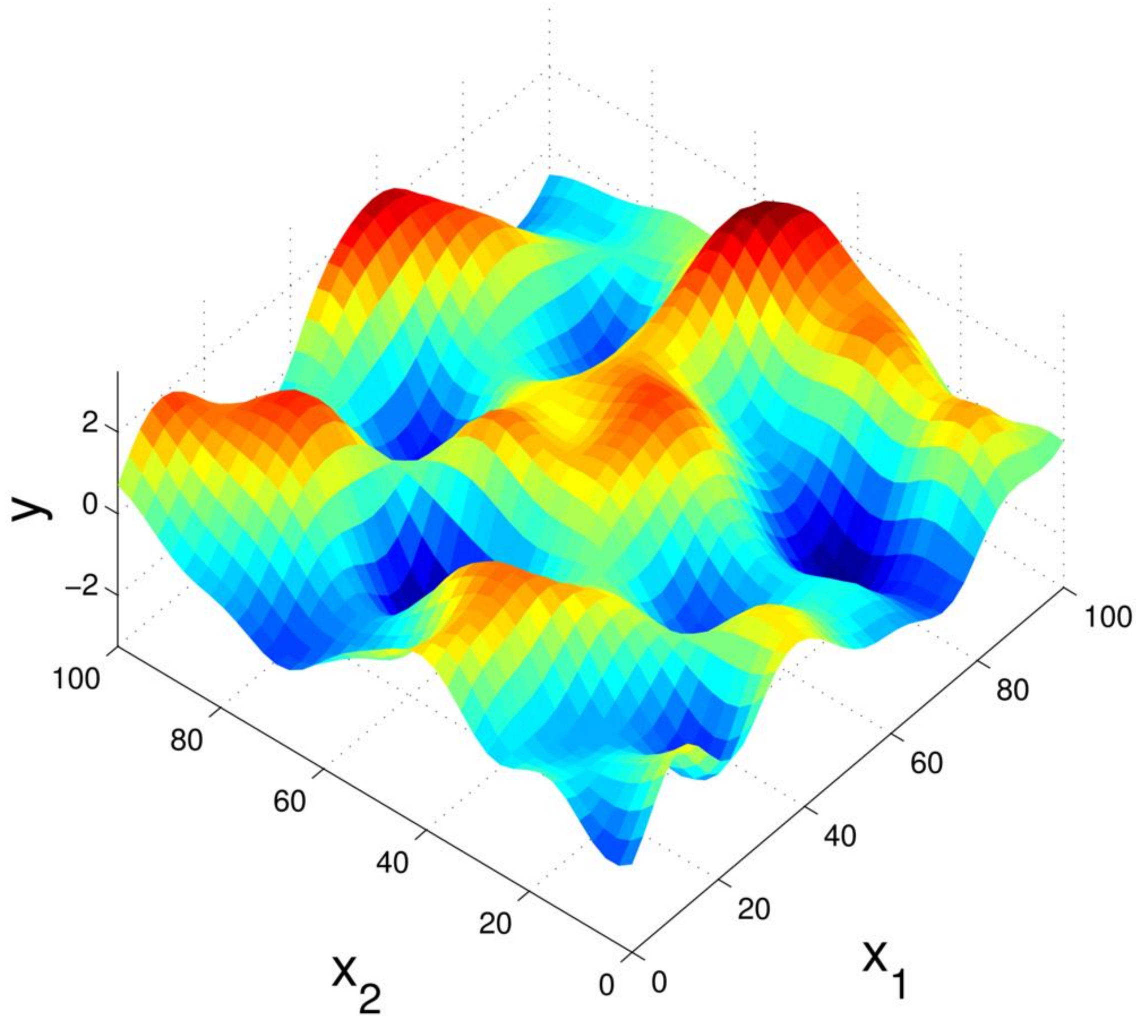
Higher Dimensional Input Spaces



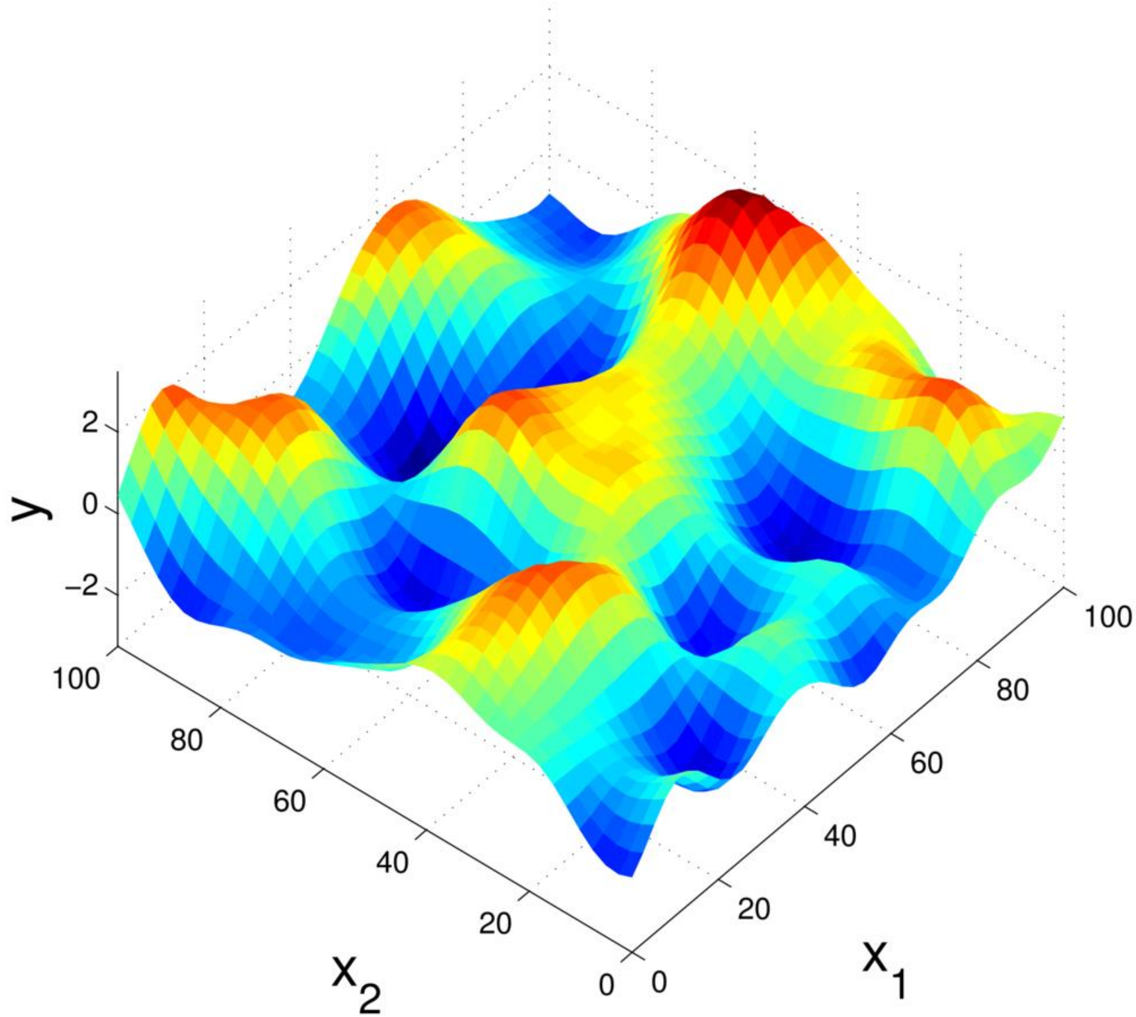
Higher Dimensional Input Spaces



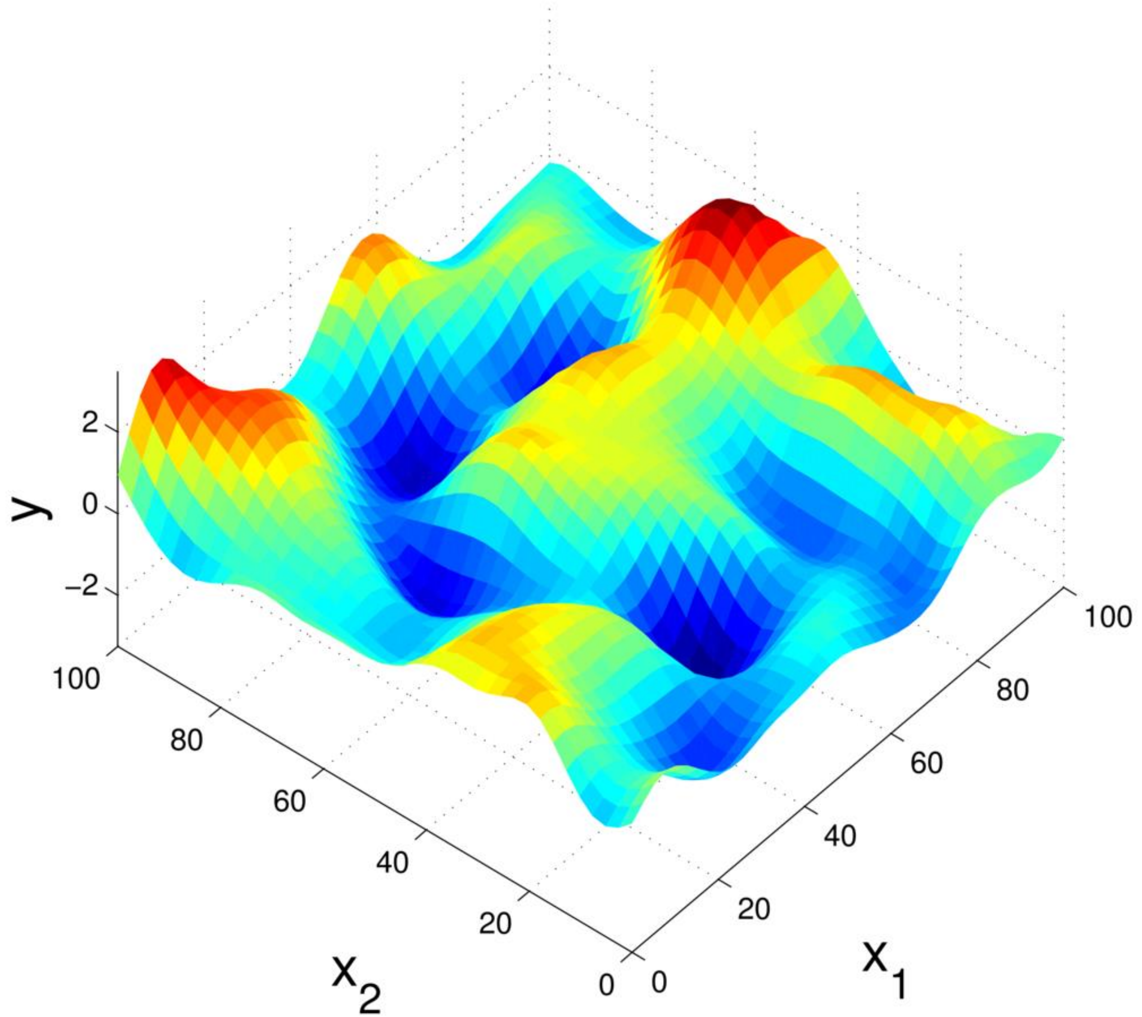
Higher Dimensional Input Spaces



Higher Dimensional Input Spaces



Higher Dimensional Input Spaces



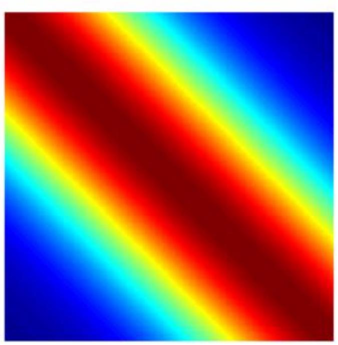
What effect do the Hyper-parameter have?

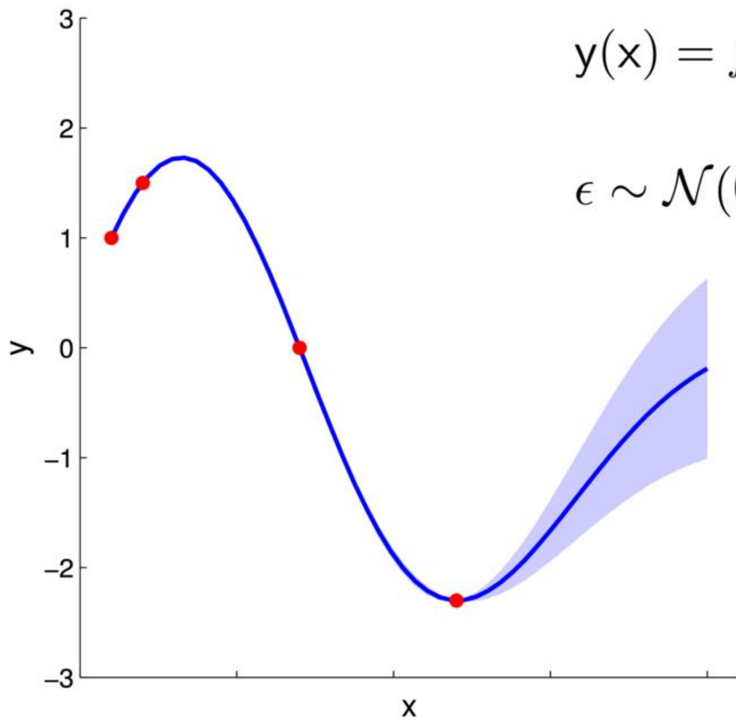
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the Hyper-parameter have?

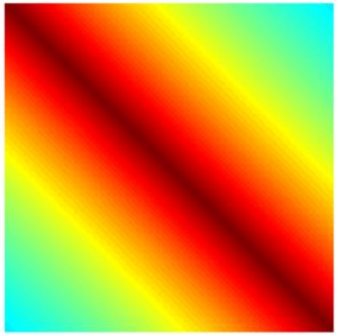
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

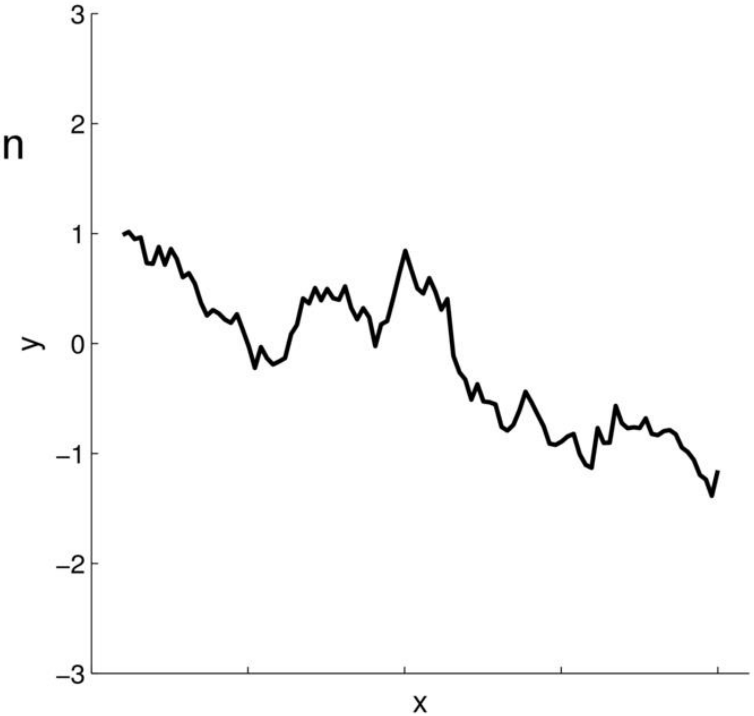
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





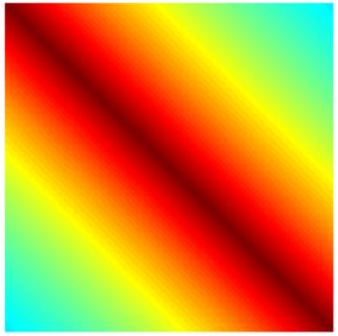
What effect do the Hyper-parameter have?

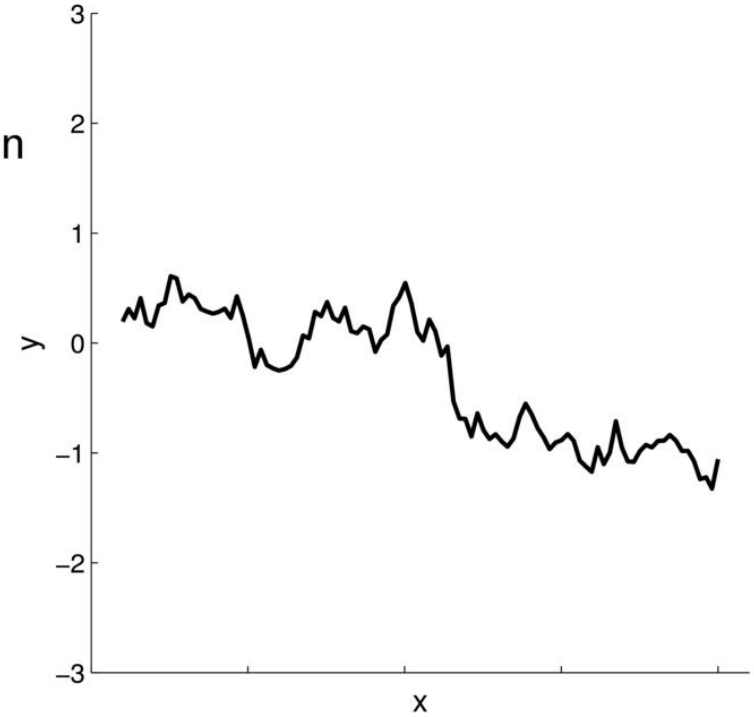
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



What effect do the Hyper-parameter have?

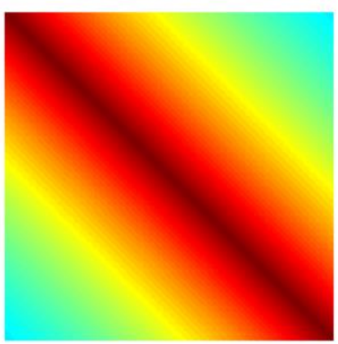
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

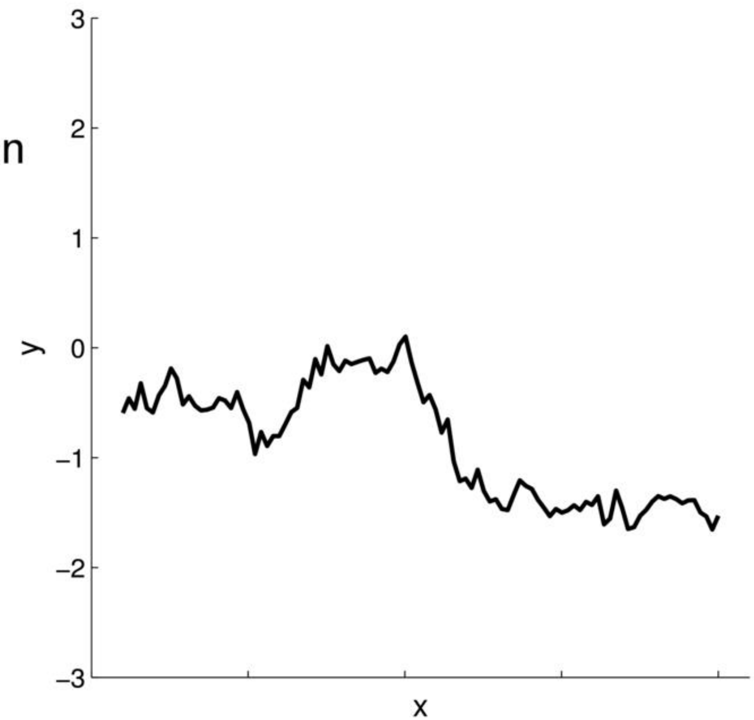
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect do the Hyper-parameter have?

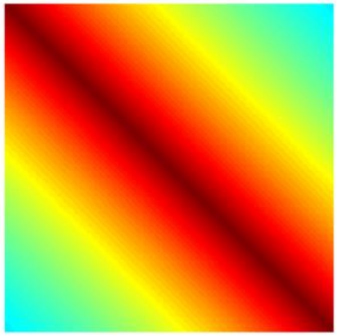
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

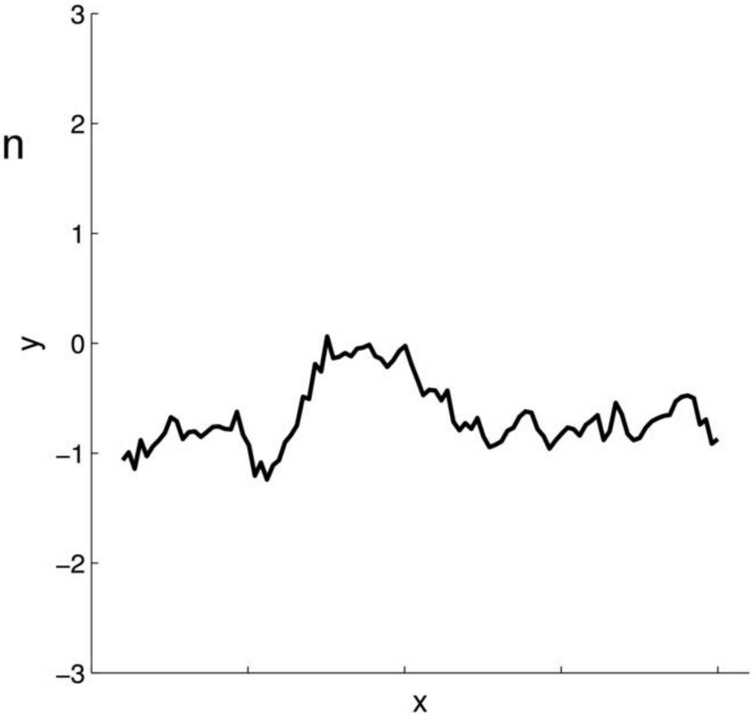
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect do the Hyper-parameter have?

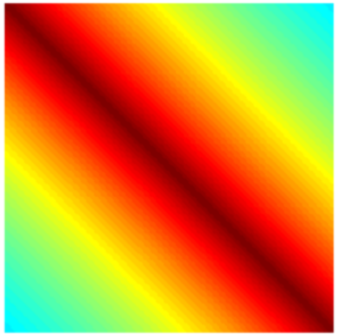
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

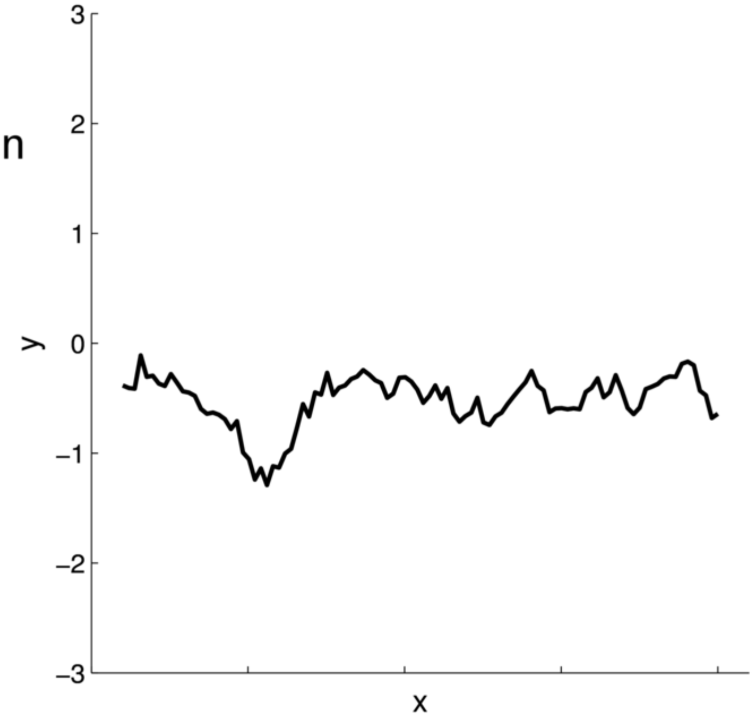
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect do the Hyper-parameter have?

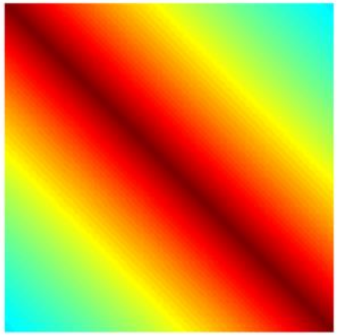
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

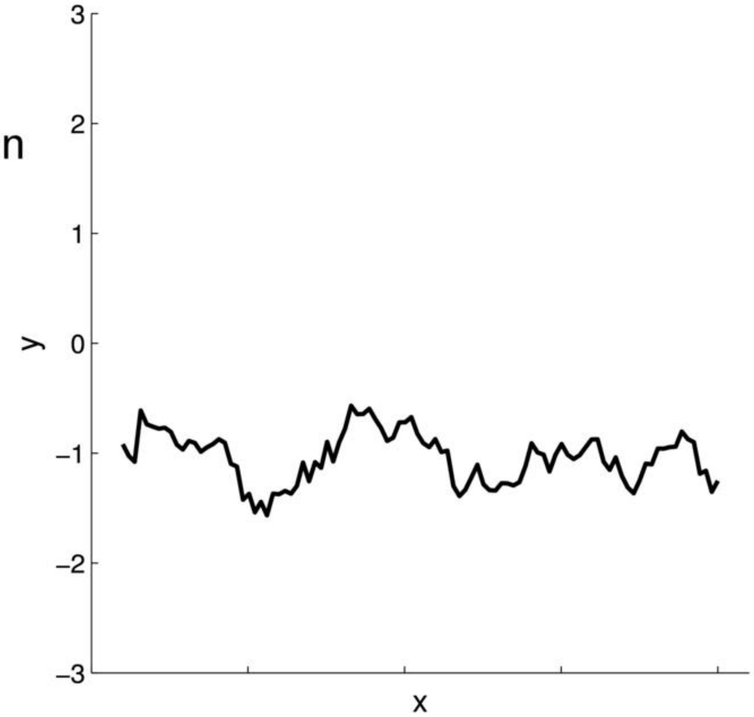
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





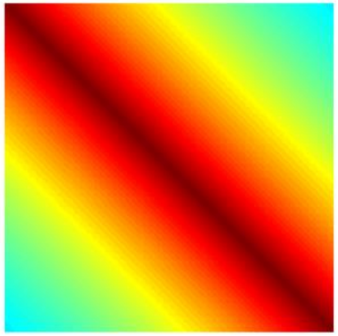
What effect do the Hyper-parameter have?

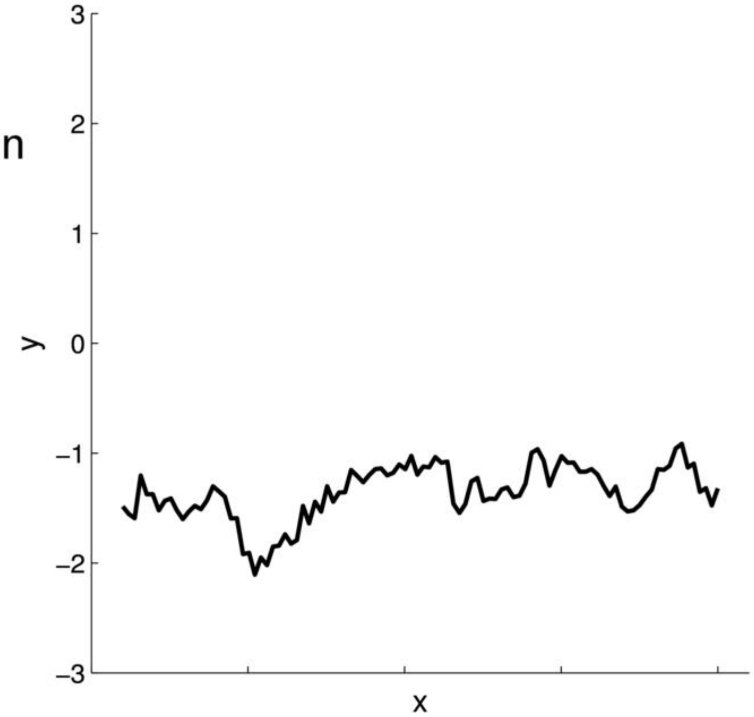
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



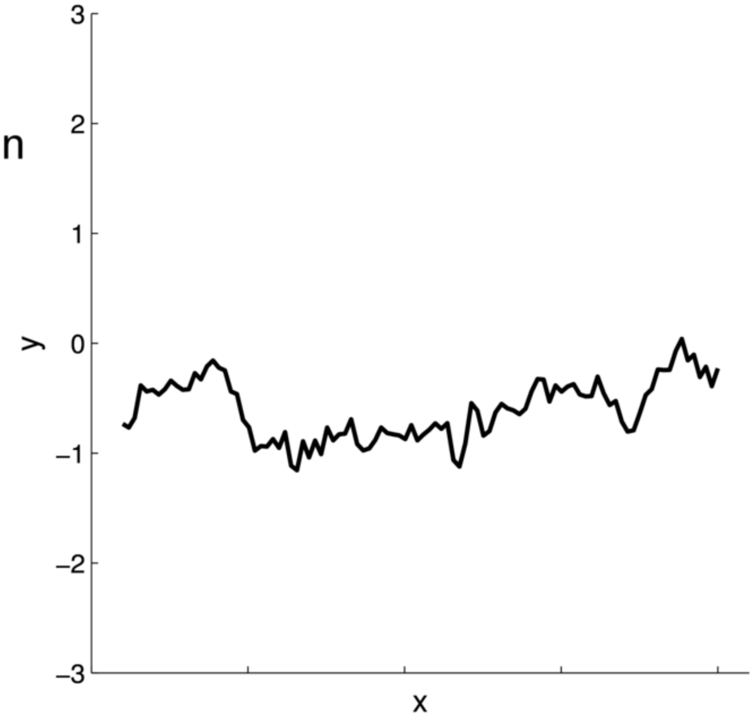
What effect do the Hyper-parameter have?

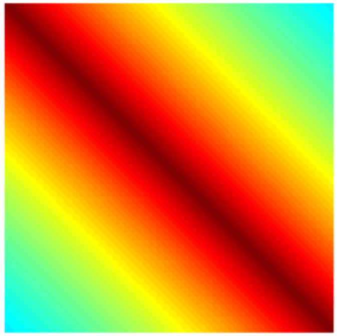
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck



$\Sigma =$ 

What effect do the Hyper-parameter have?

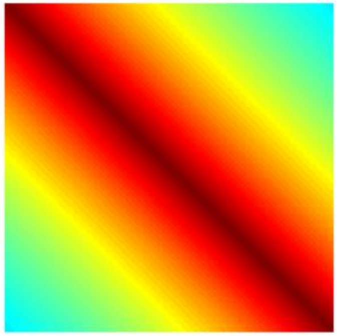
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

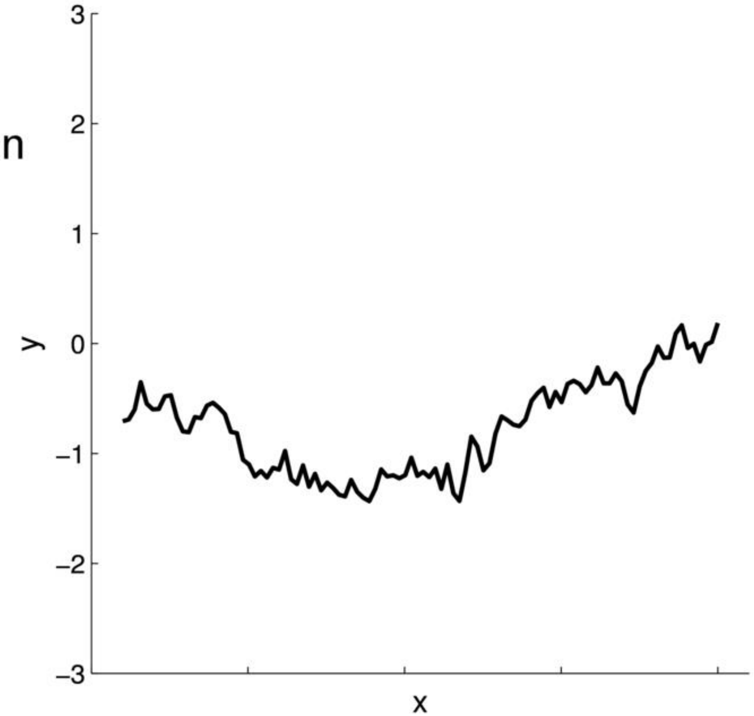
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





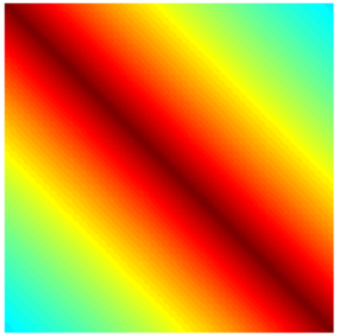
What effect do the Hyper-parameter have?

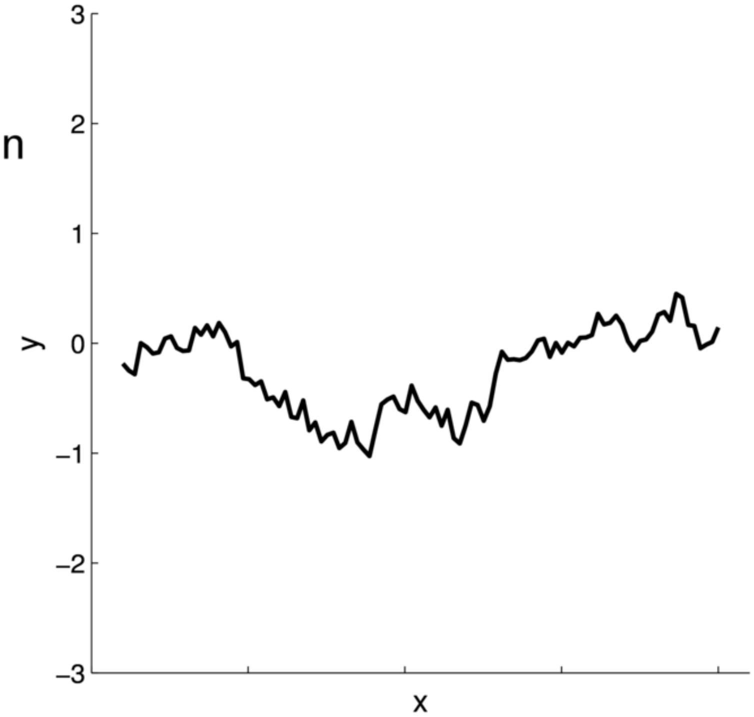
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



What effect do the Hyper-parameter have?

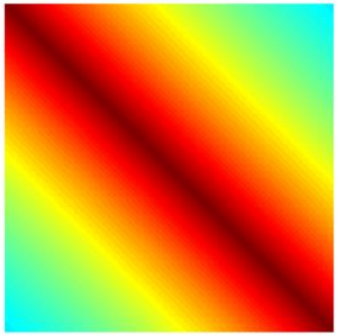
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

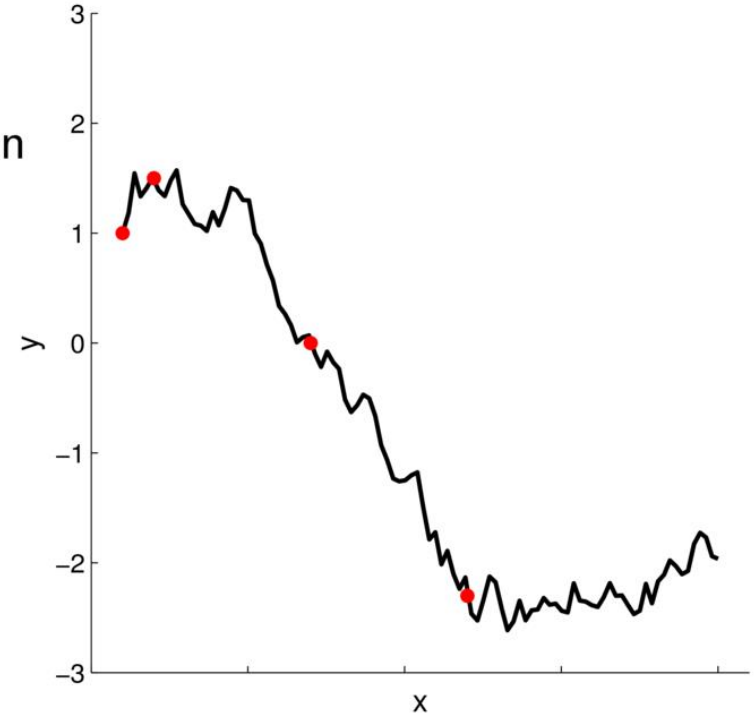
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





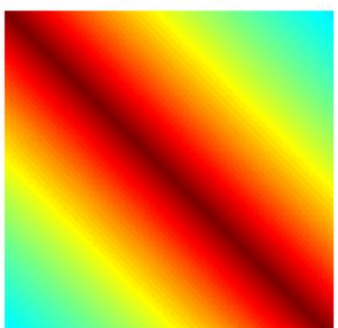
What effect do the Hyper-parameter have?

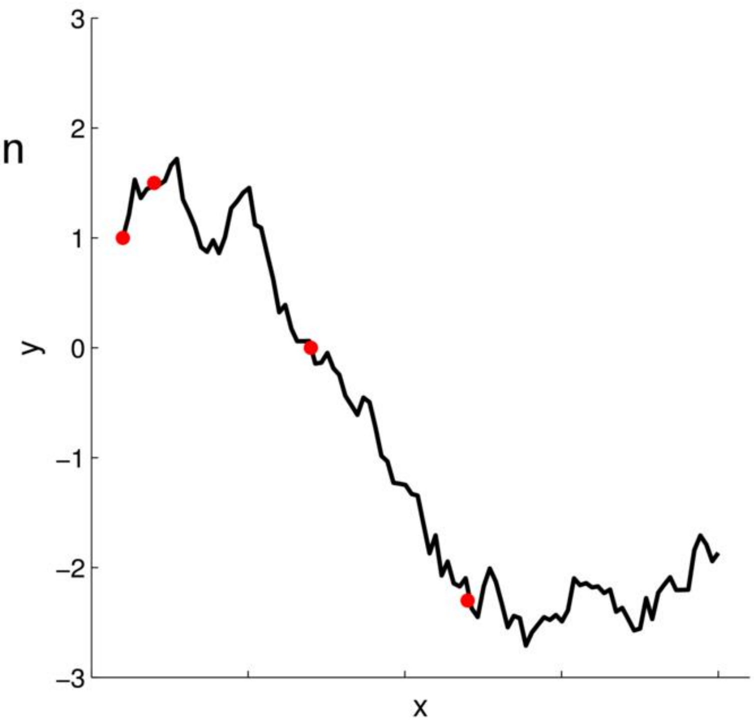
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



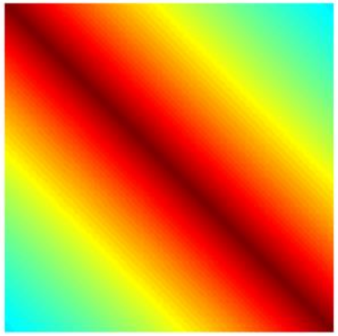
What effect do the Hyper-parameter have?

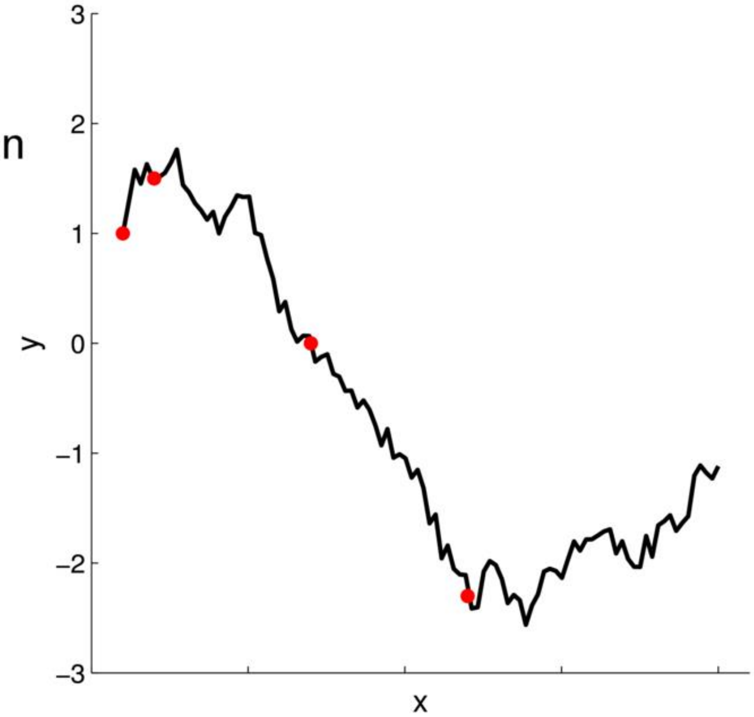
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



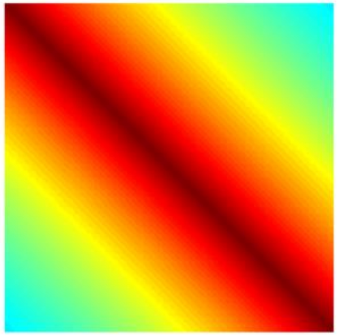
What effect do the Hyper-parameter have?

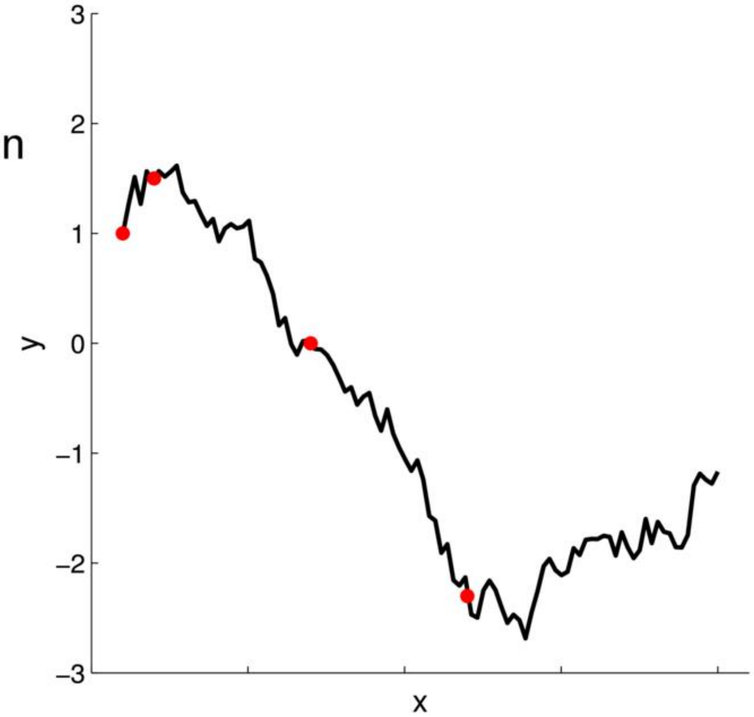
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



What effect do the Hyper-parameter have?

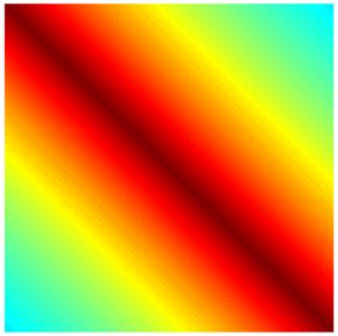
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

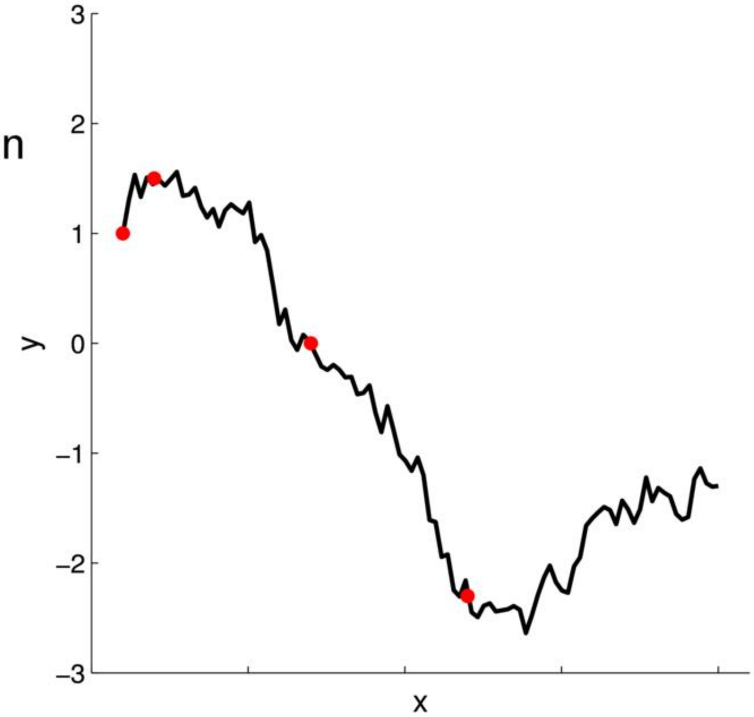
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect do the Hyper-parameter have?

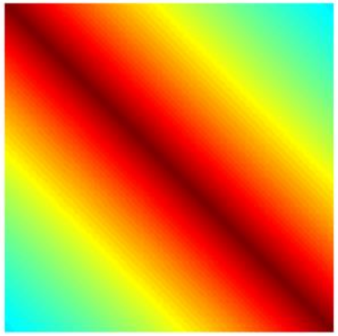
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

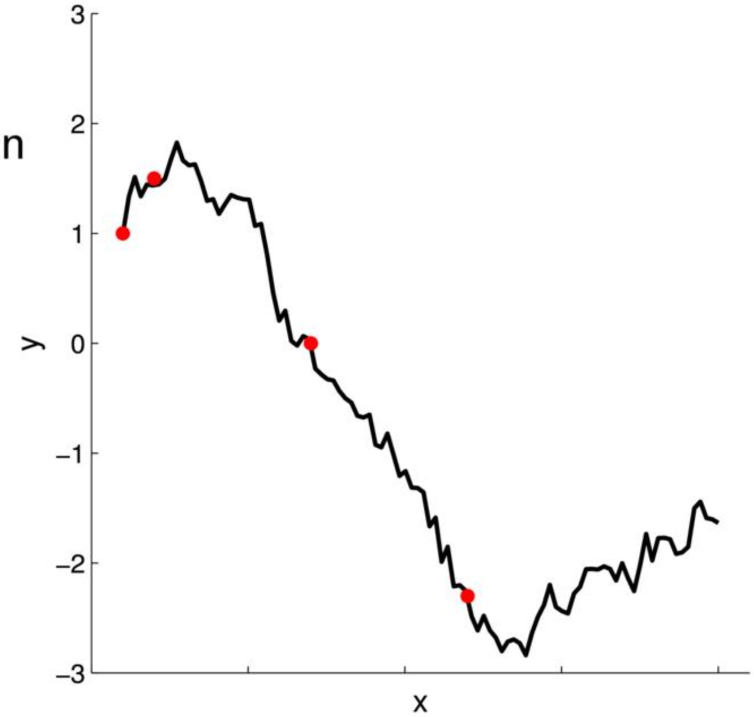
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect do the Hyper-parameter have?

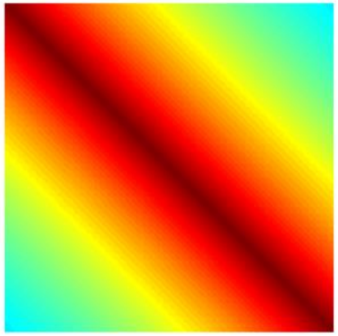
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

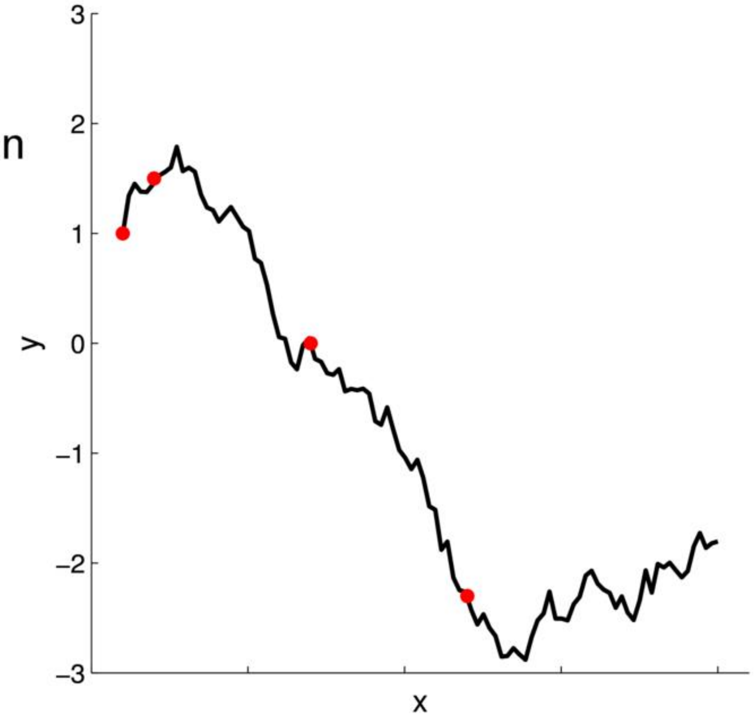
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect do the Hyper-parameter have?

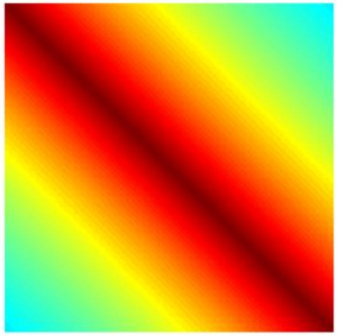
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

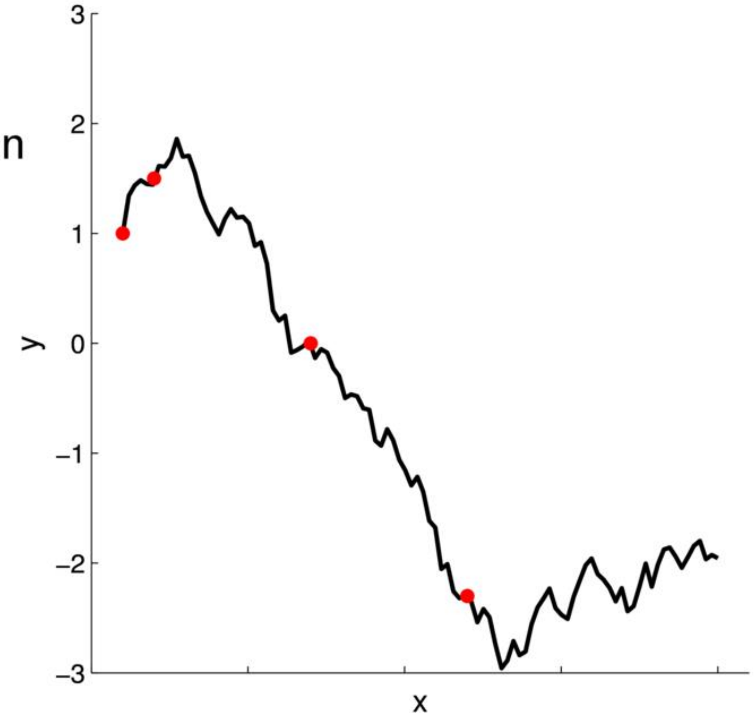
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect do the Hyper-parameter have?

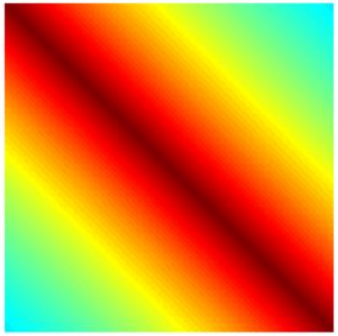
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

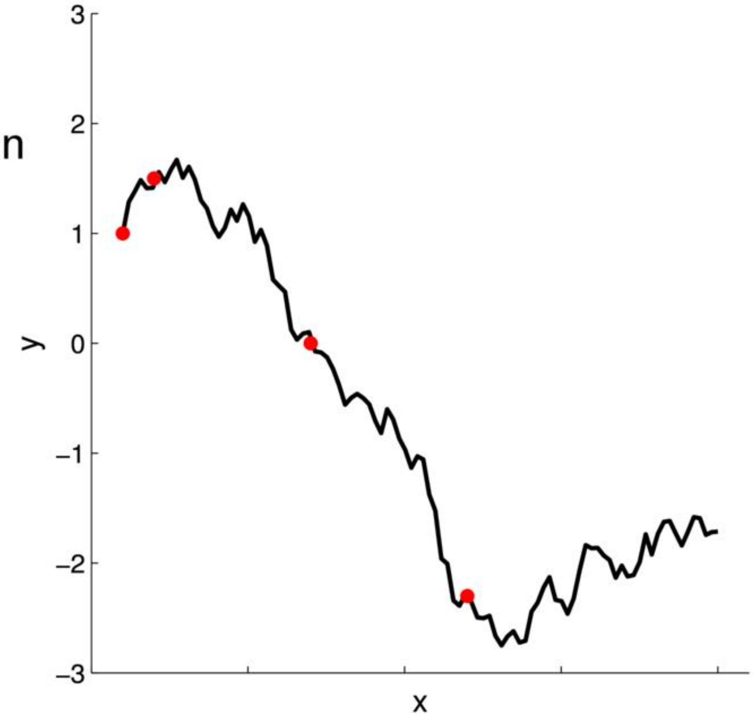
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





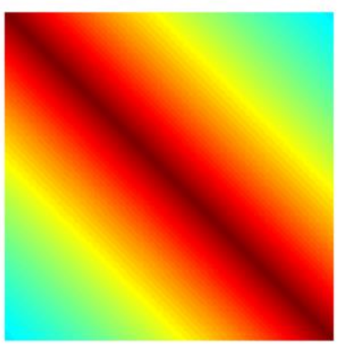
What effect do the Hyper-parameter have?

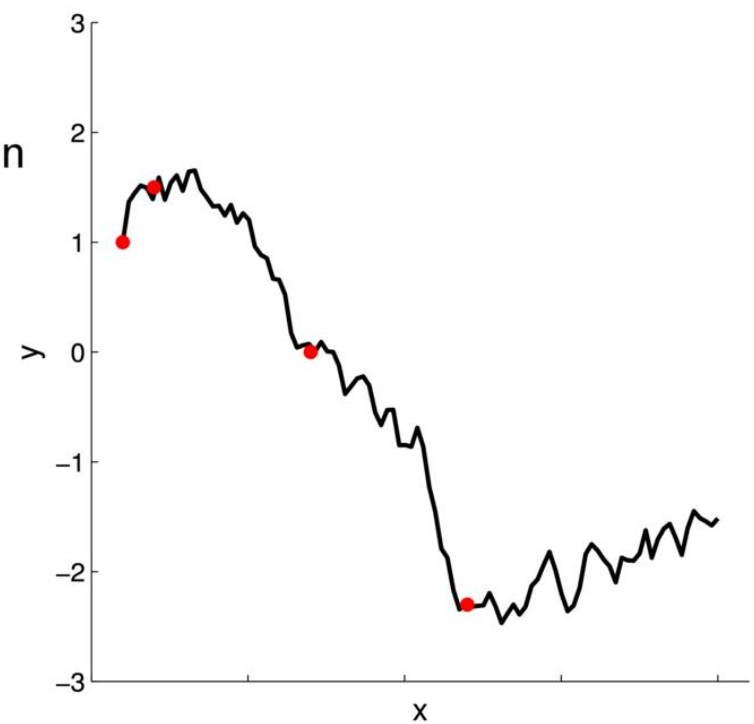
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



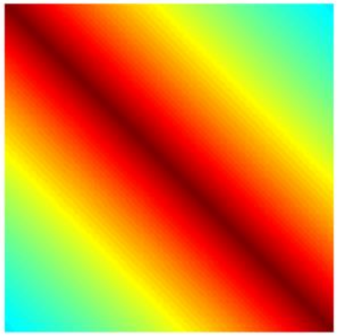
What effect do the Hyper-parameter have?

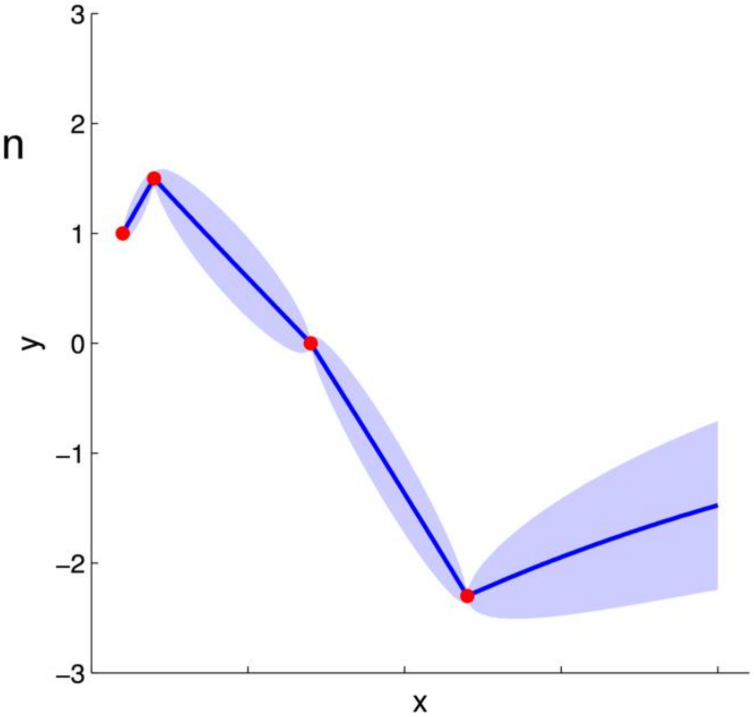
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 

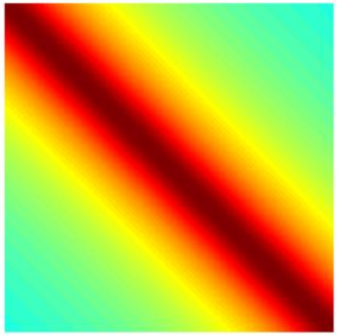


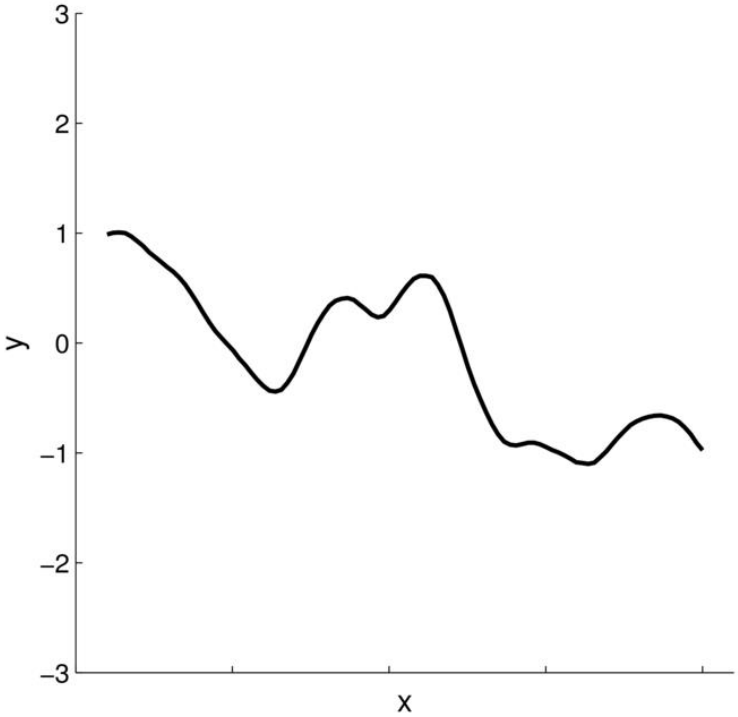
What effect do the Hyper-parameter have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

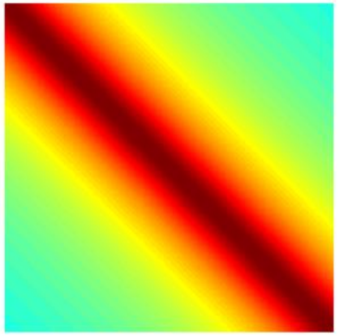


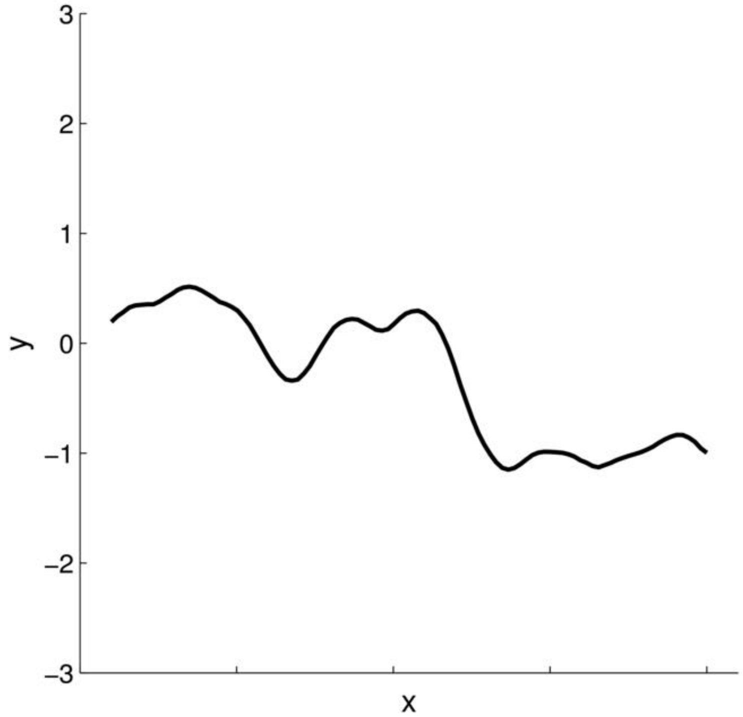


What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

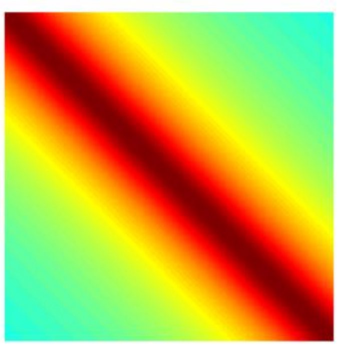


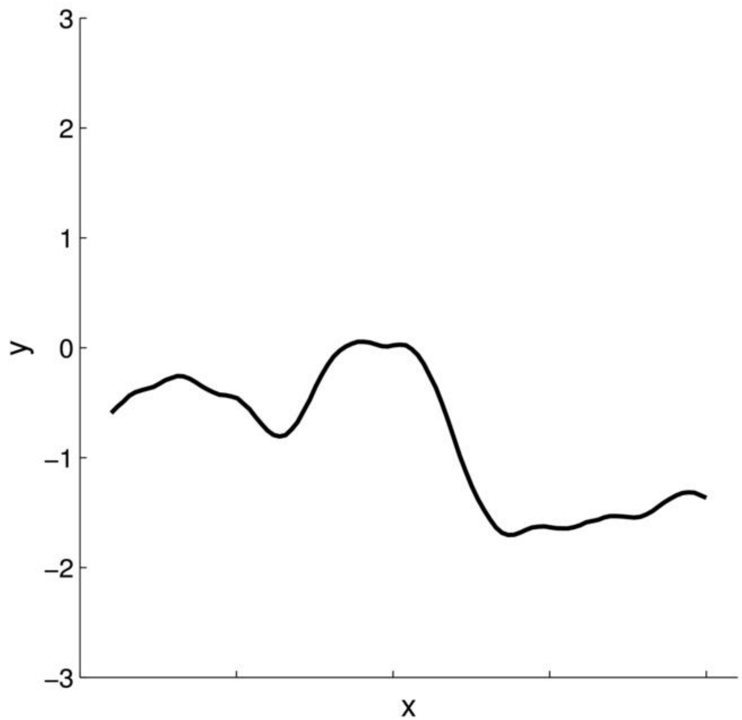
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

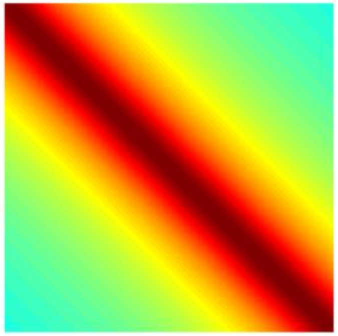


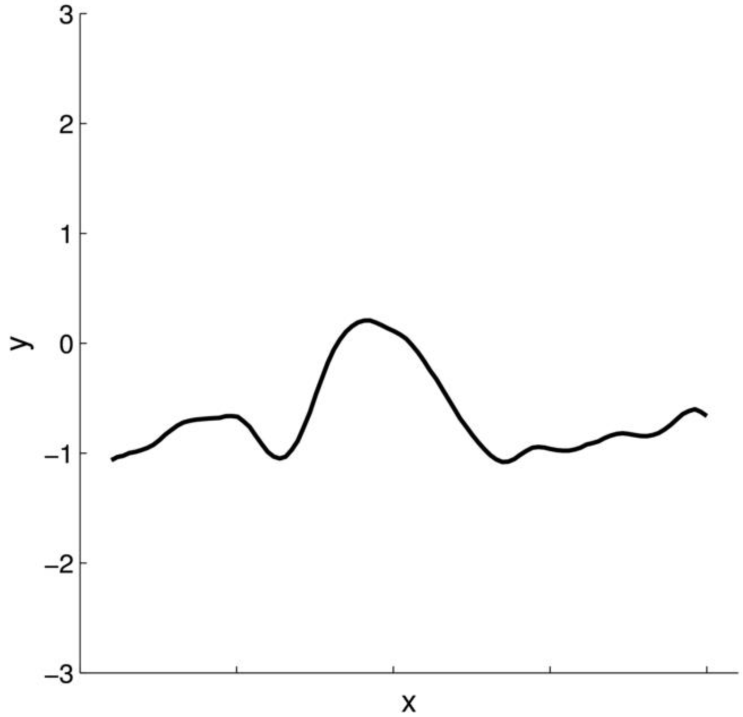


What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

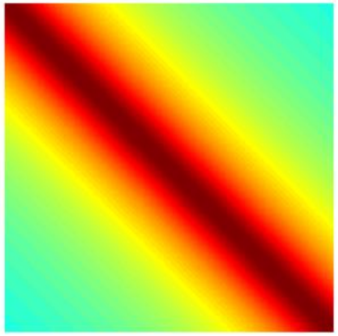
$\Sigma =$ 

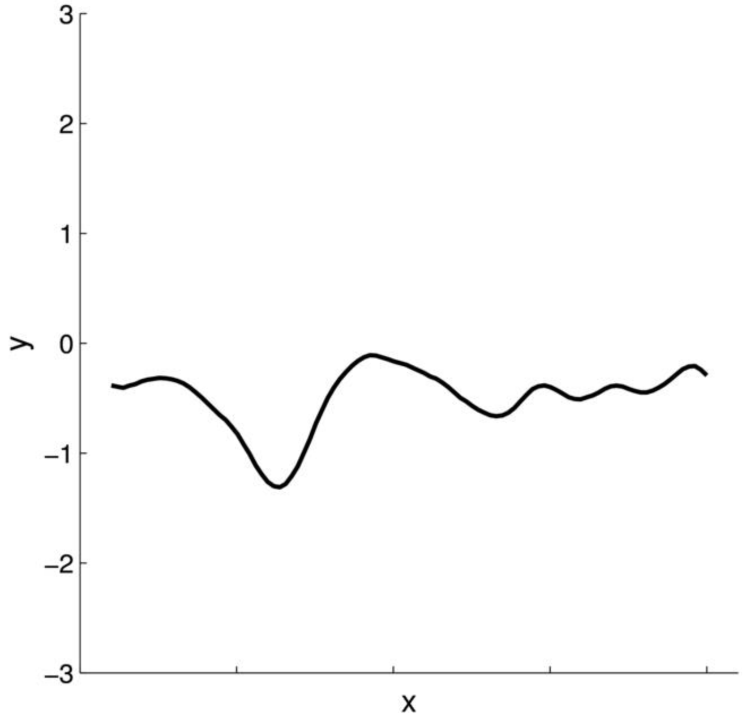


What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

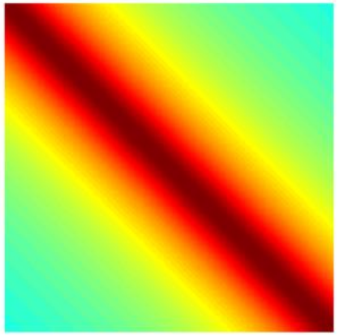
$\Sigma =$ 

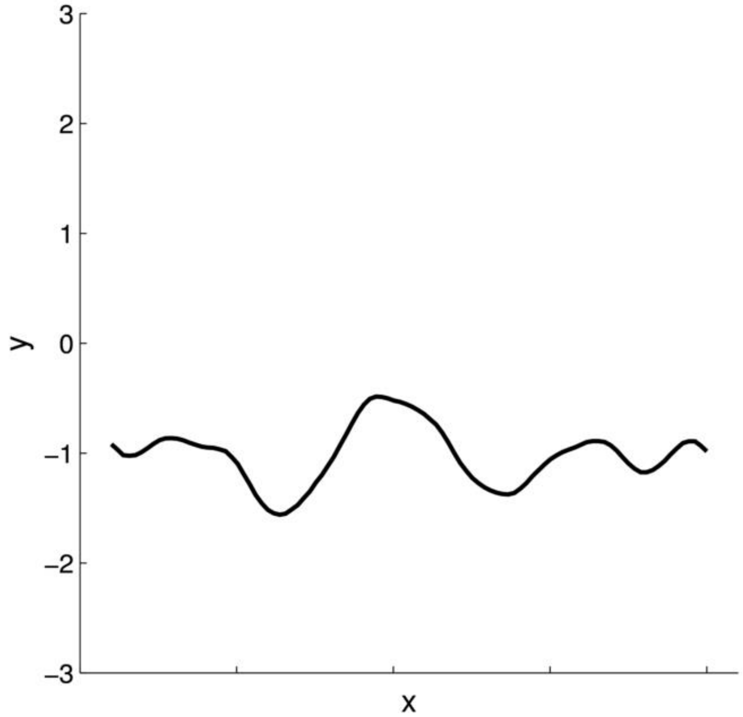


What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

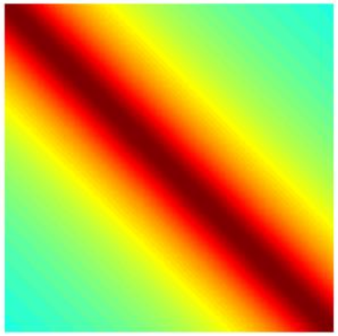


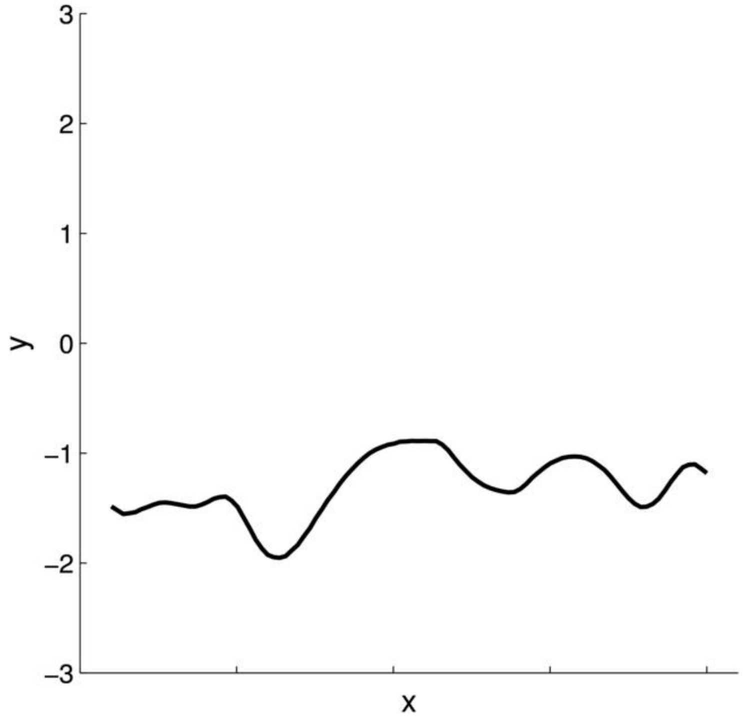
What effect do the Hyper-parameter have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

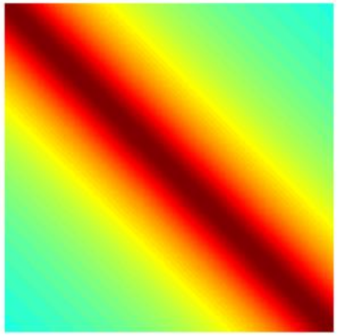


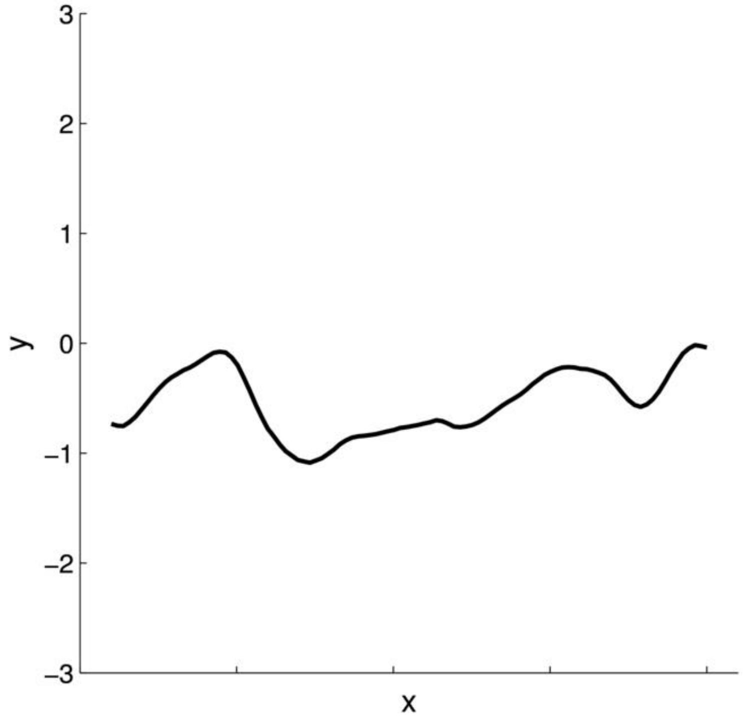


What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

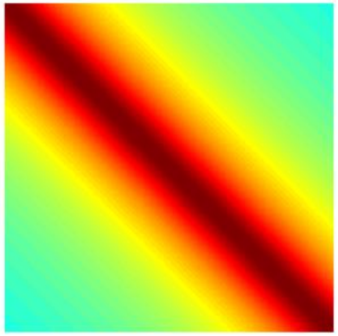
$\Sigma =$ 

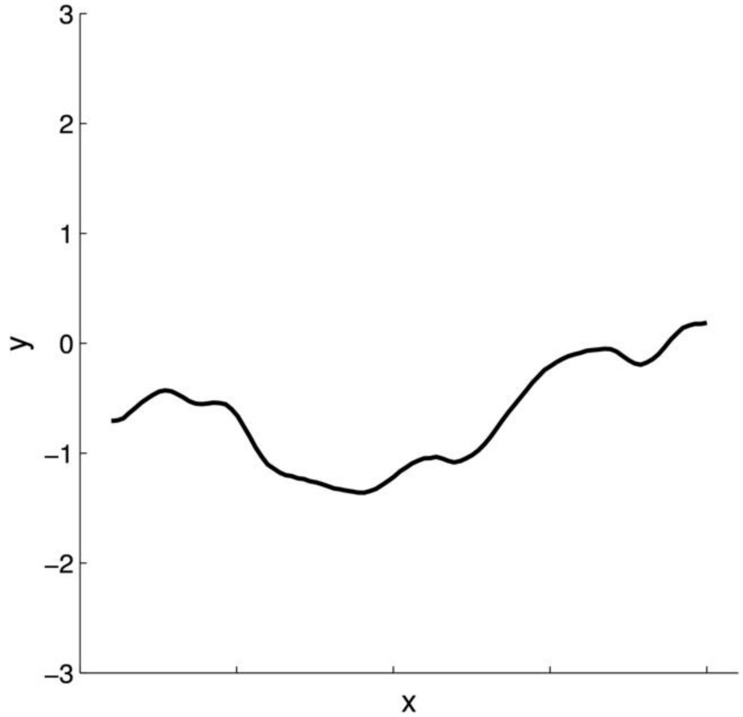


What effect do the Hyper-parameter have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

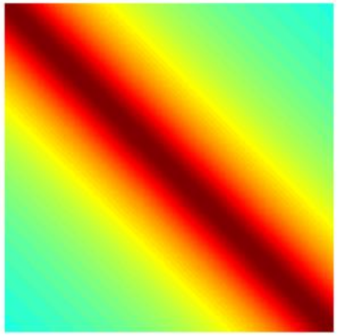
$\Sigma =$ 

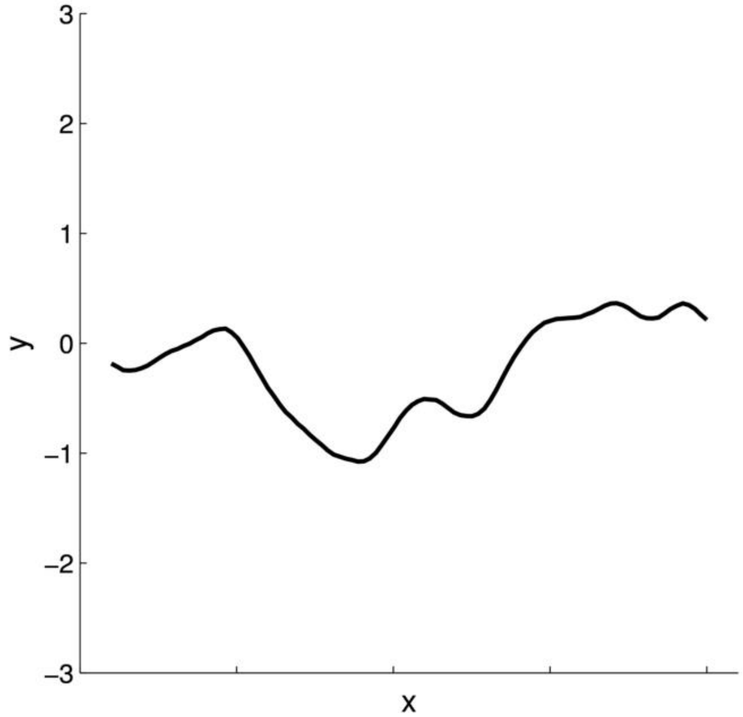


What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

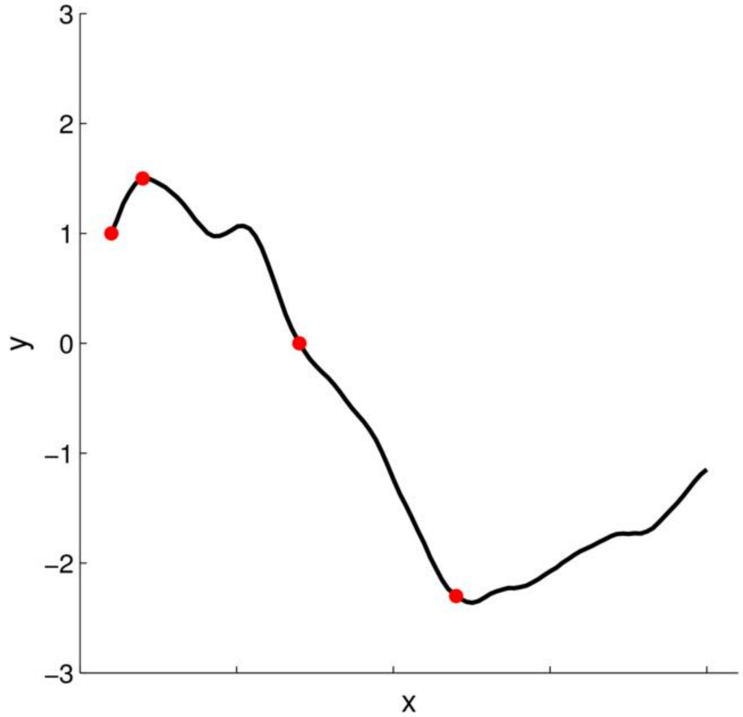
$\Sigma =$ 

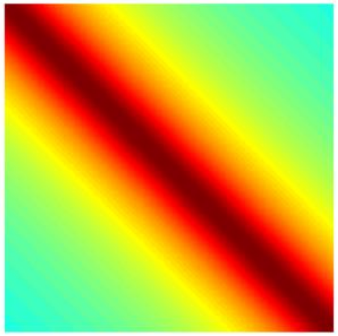


What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

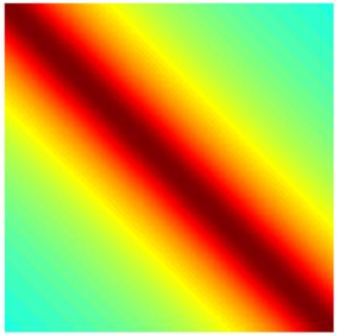


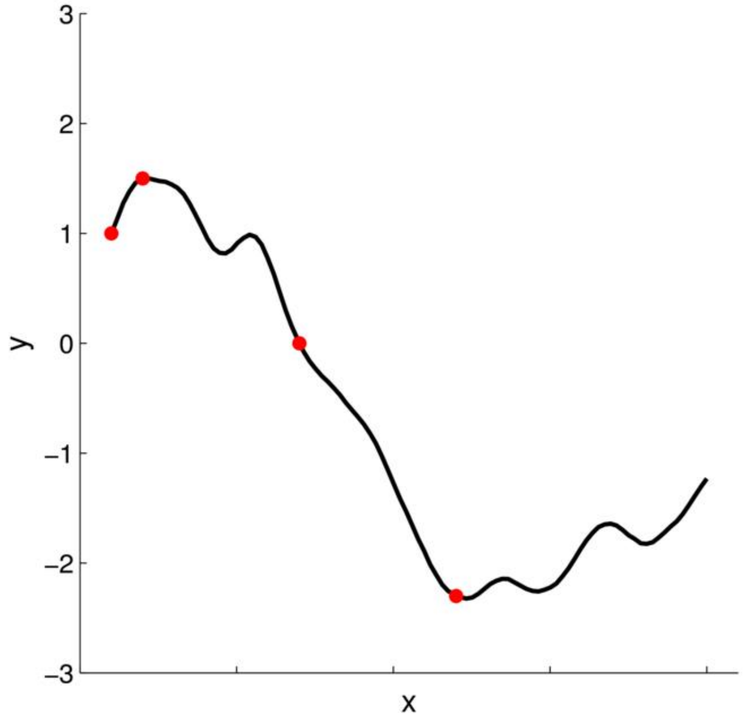
$\Sigma =$ 

What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

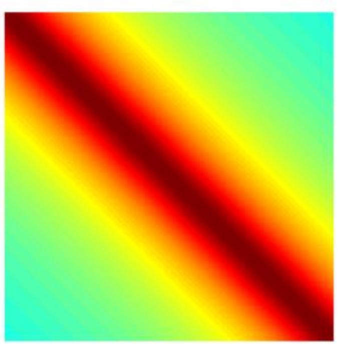


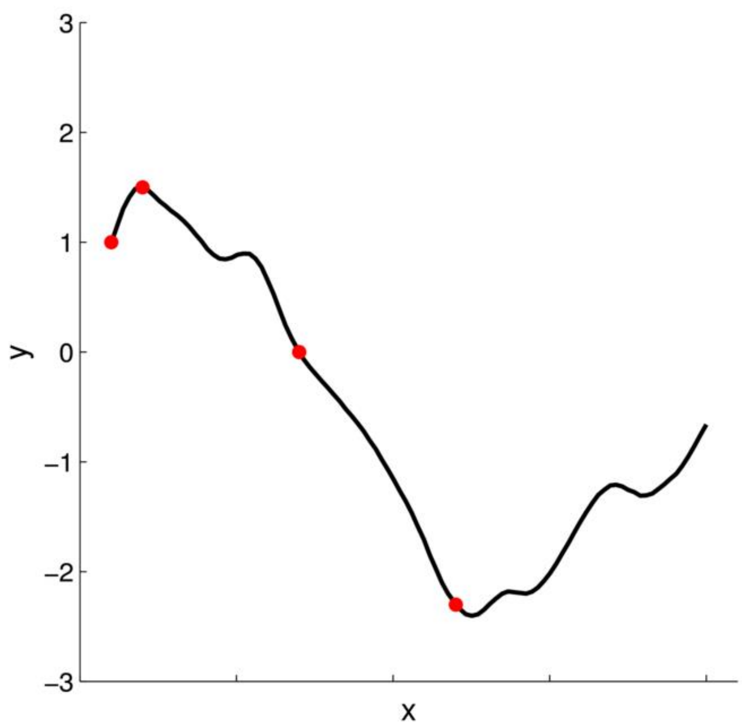
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

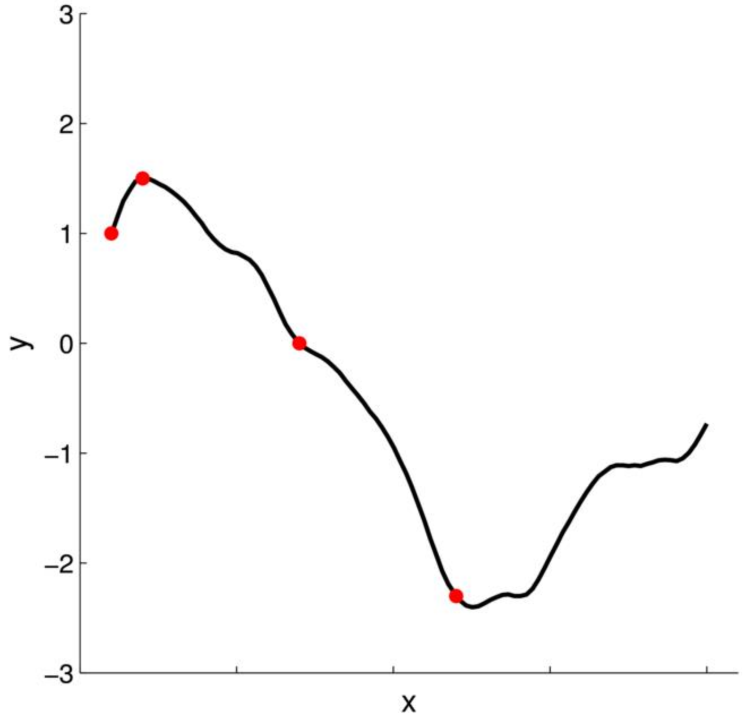




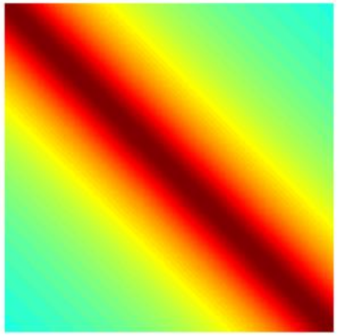
What effect do the Hyper-parameter have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic



$\Sigma =$

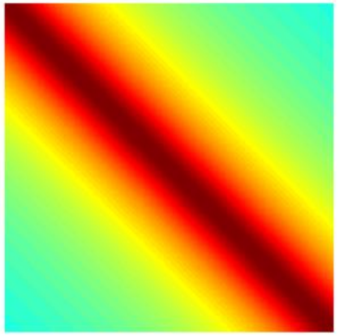


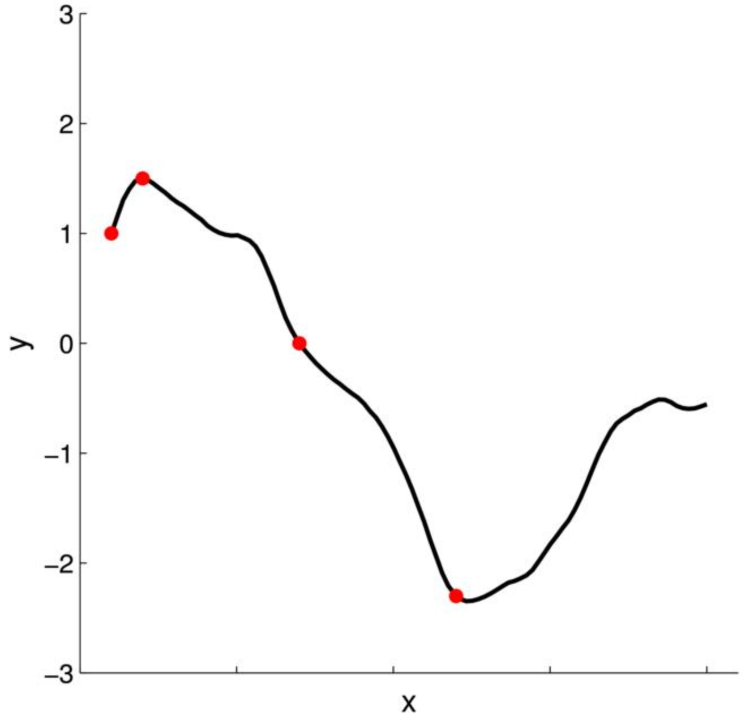
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$



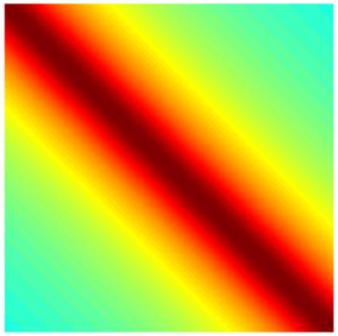


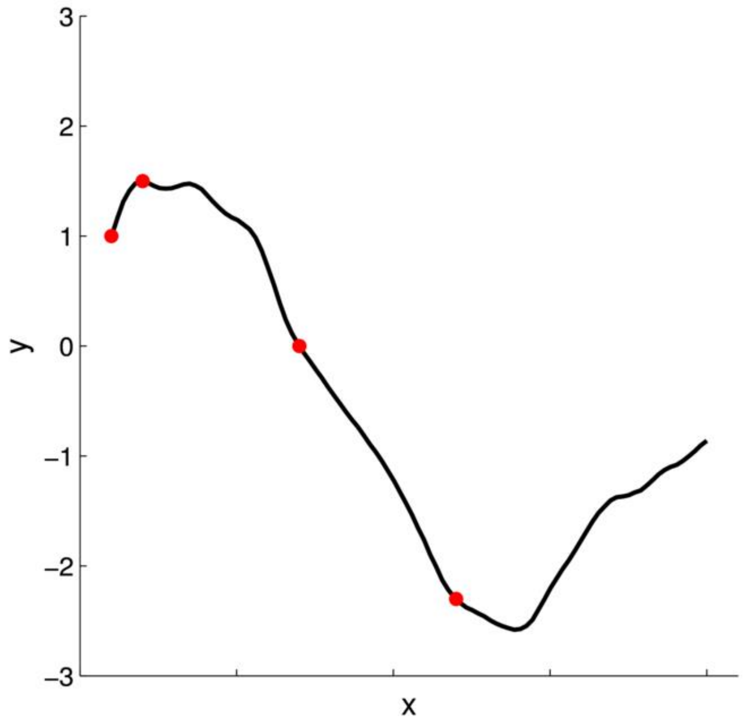
What effect do the Hyper-parameter have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$



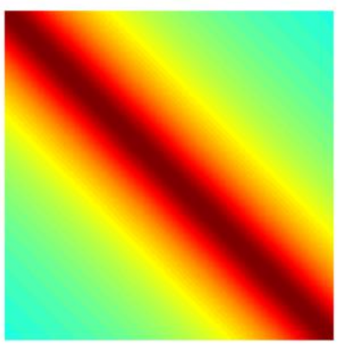


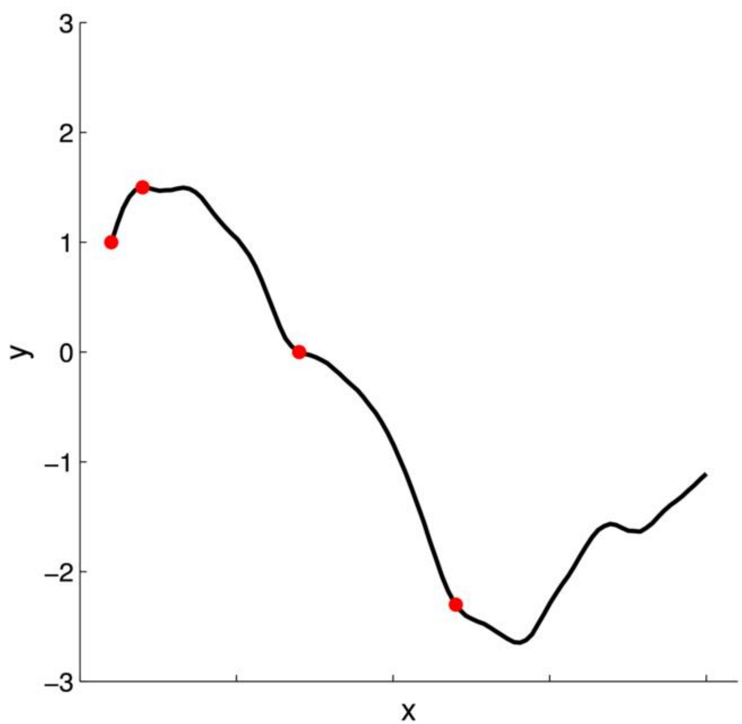
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

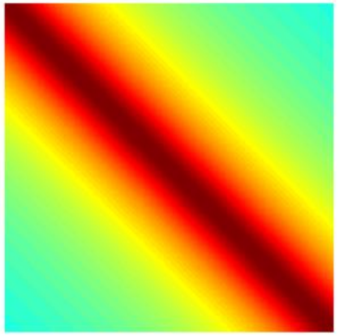


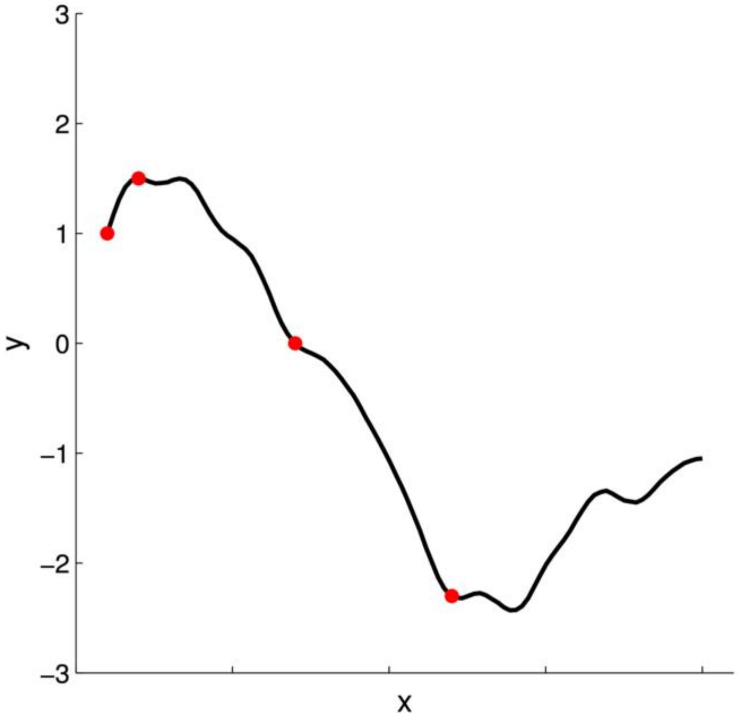


What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

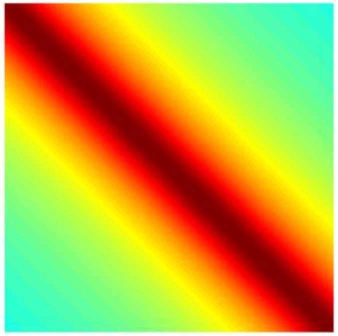


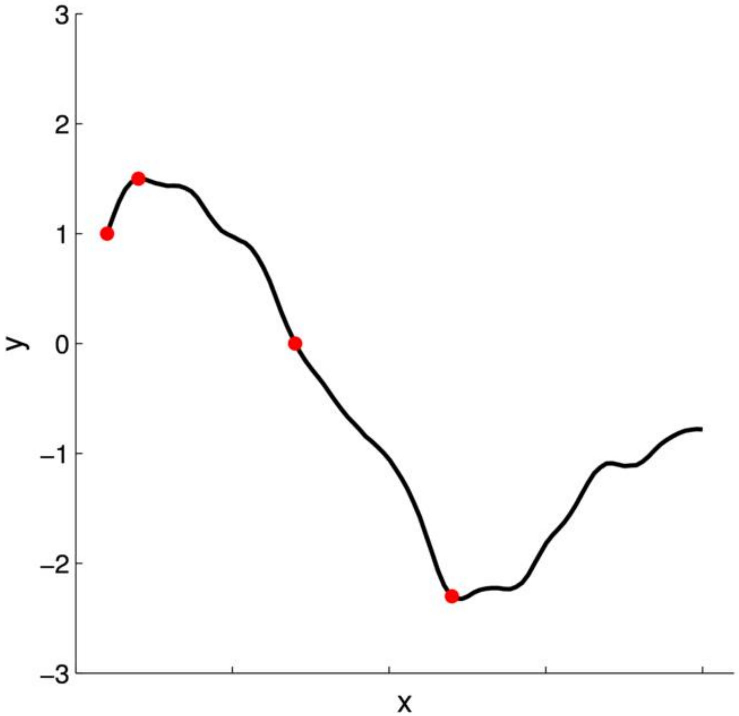
What effect do the Hyper-parameter have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

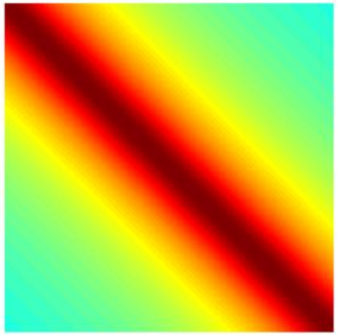


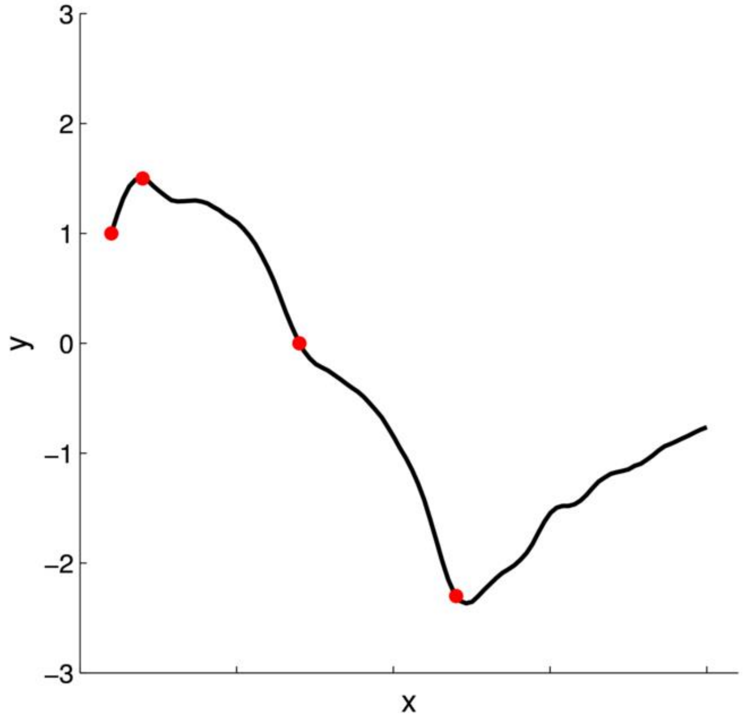


What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

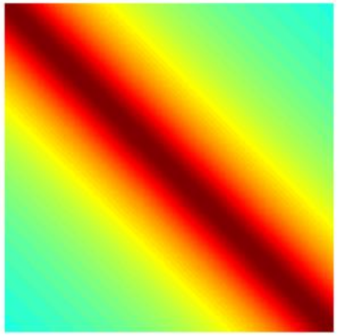
$\Sigma =$ 

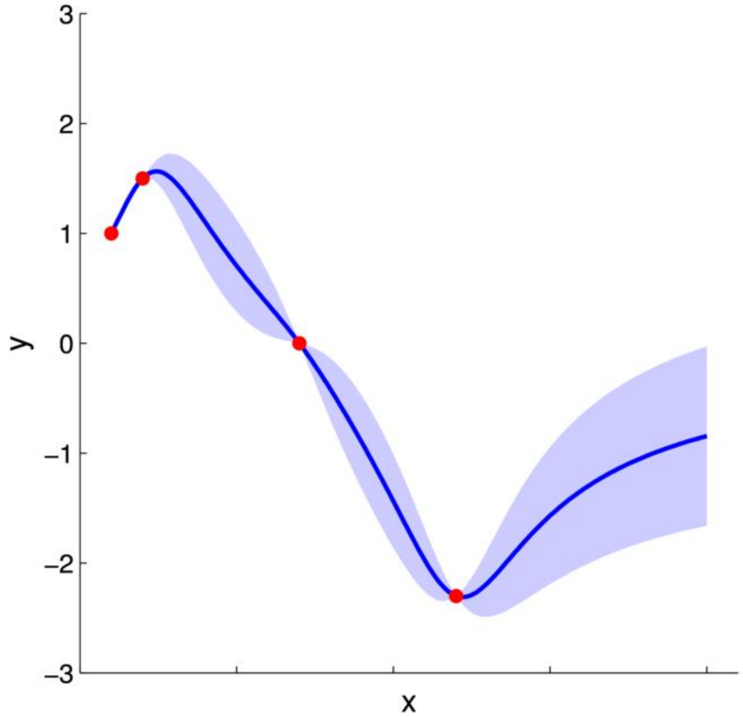


What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

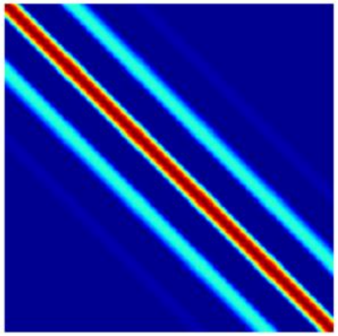


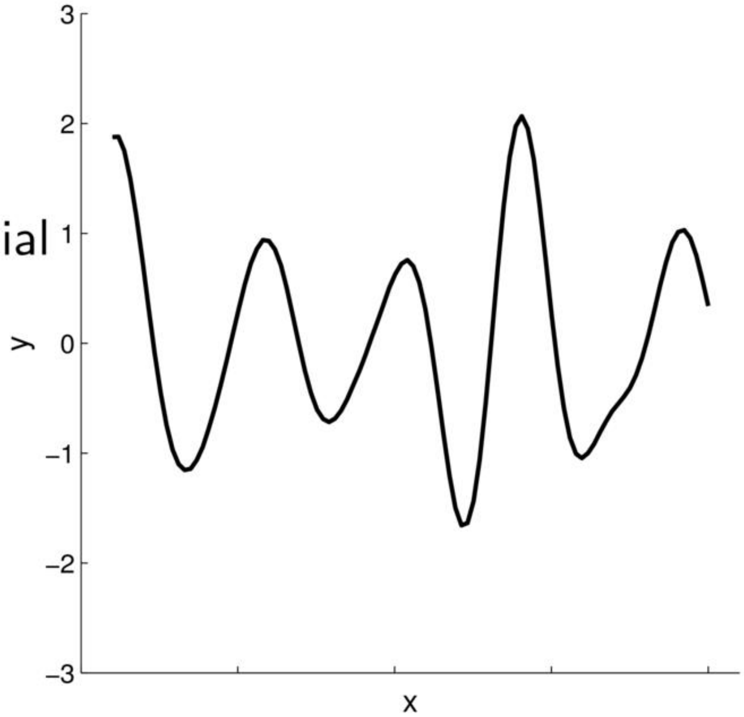
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

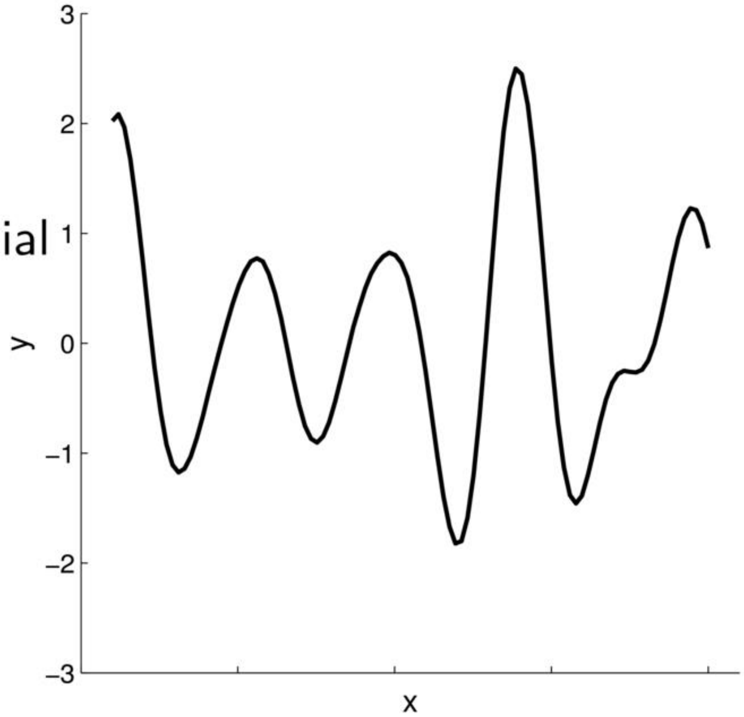


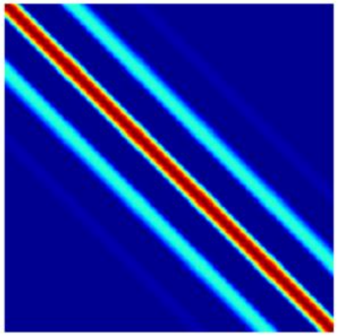
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



$\Sigma =$ 

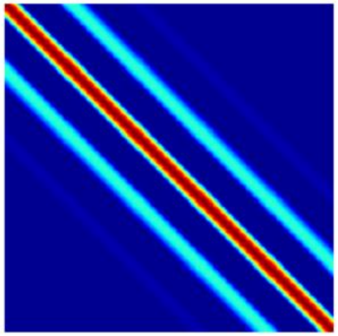
What effect do the Hyper-parameter have?

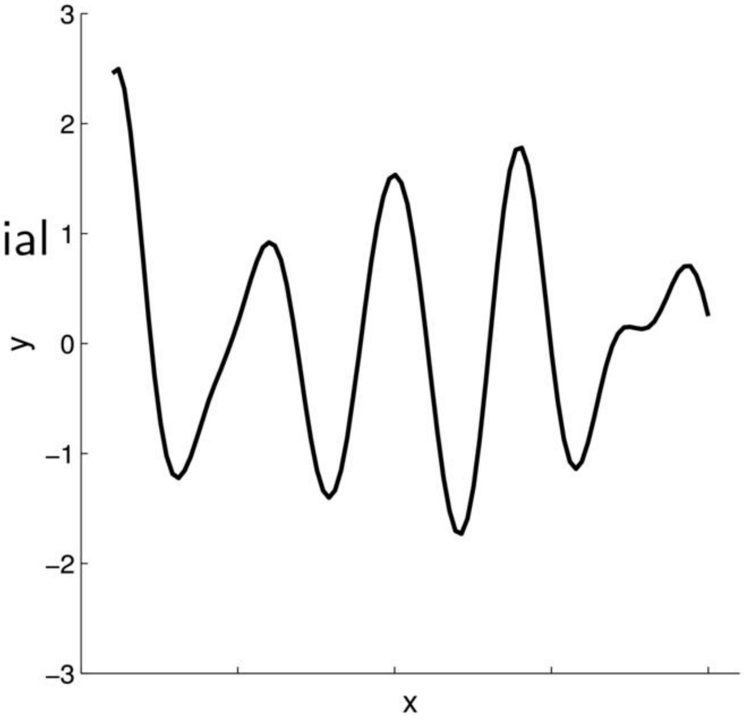
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$





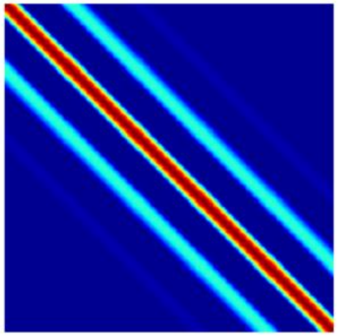
What effect do the Hyper-parameter have?

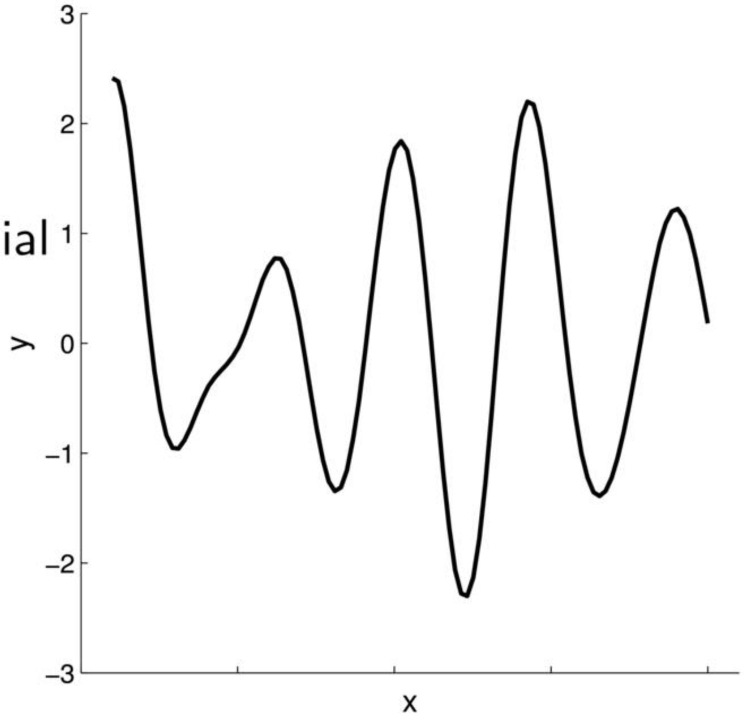
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$



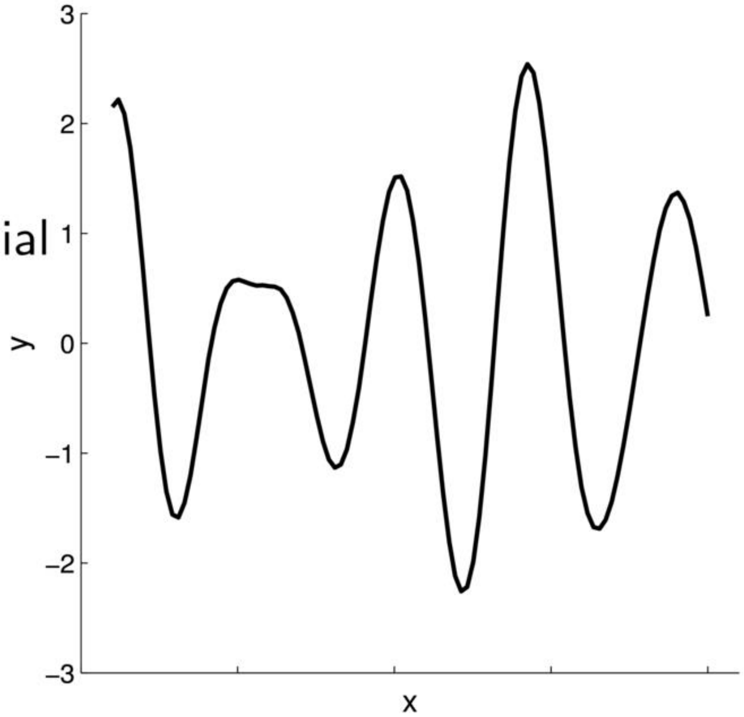


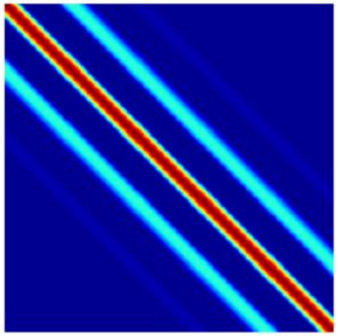
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



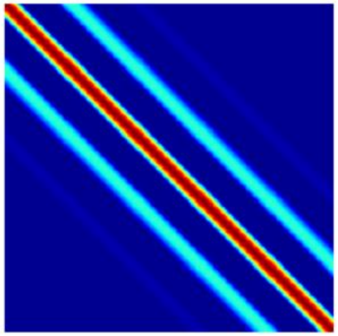
$\Sigma =$ 

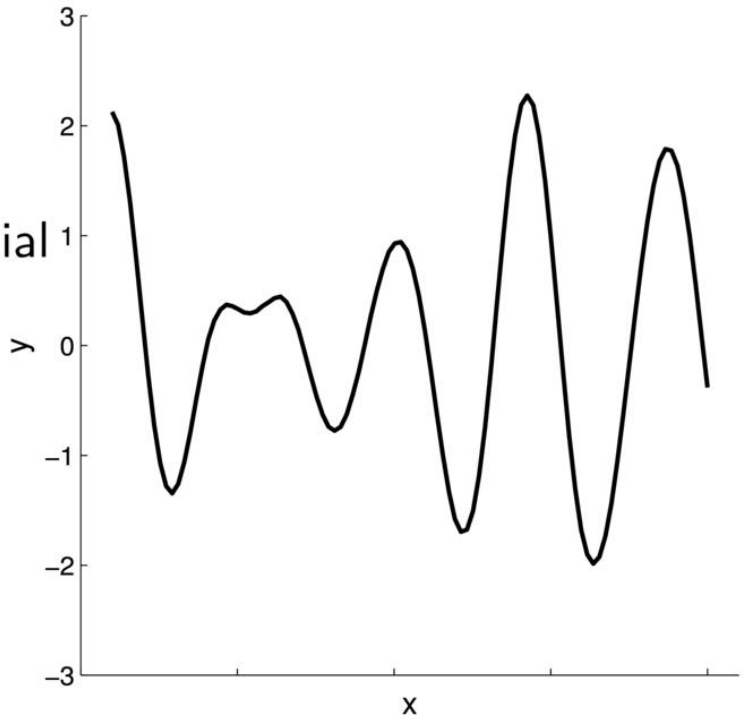
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

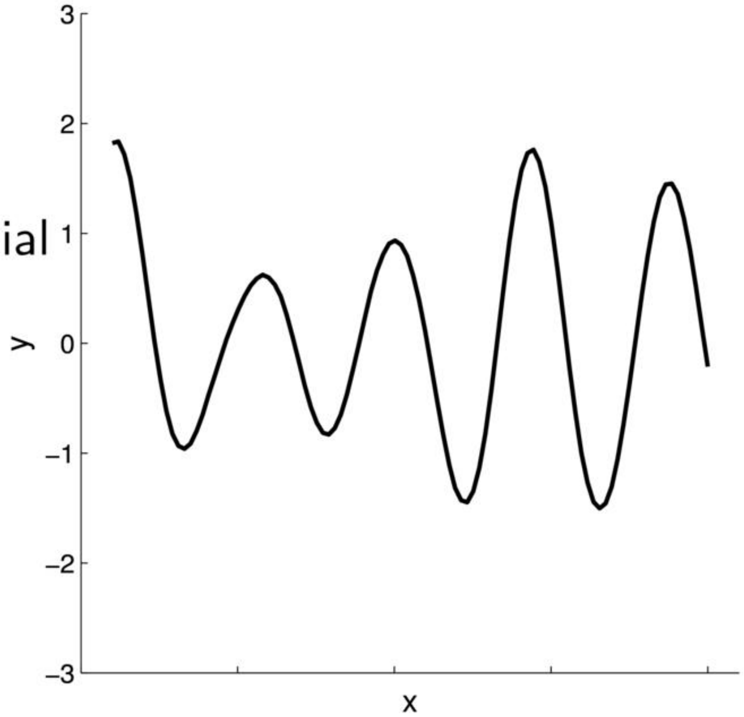


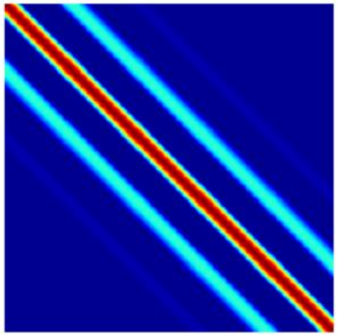
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



$\Sigma =$ 

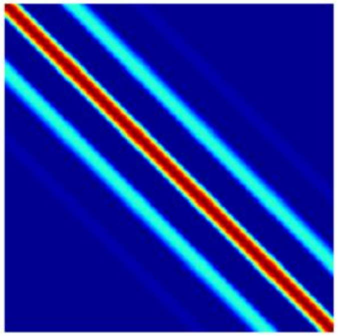
What effect do the Hyper-parameter have?

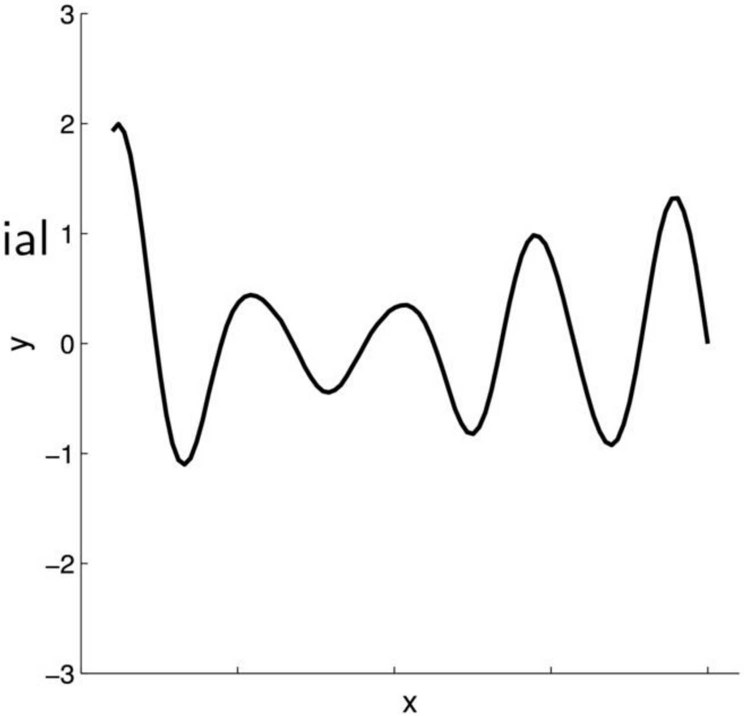
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$





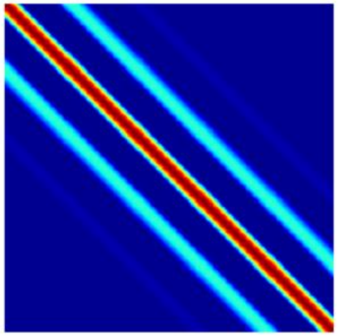
What effect do the Hyper-parameter have?

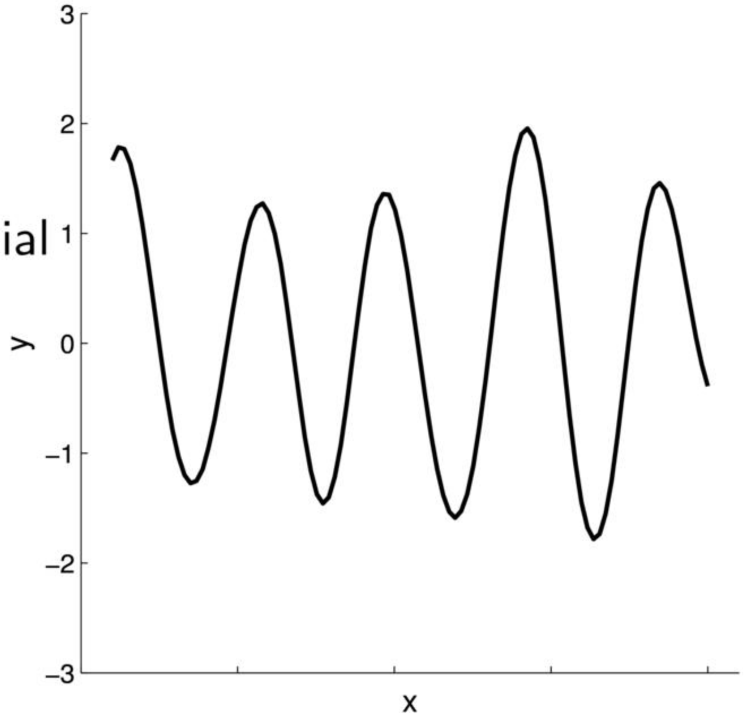
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$



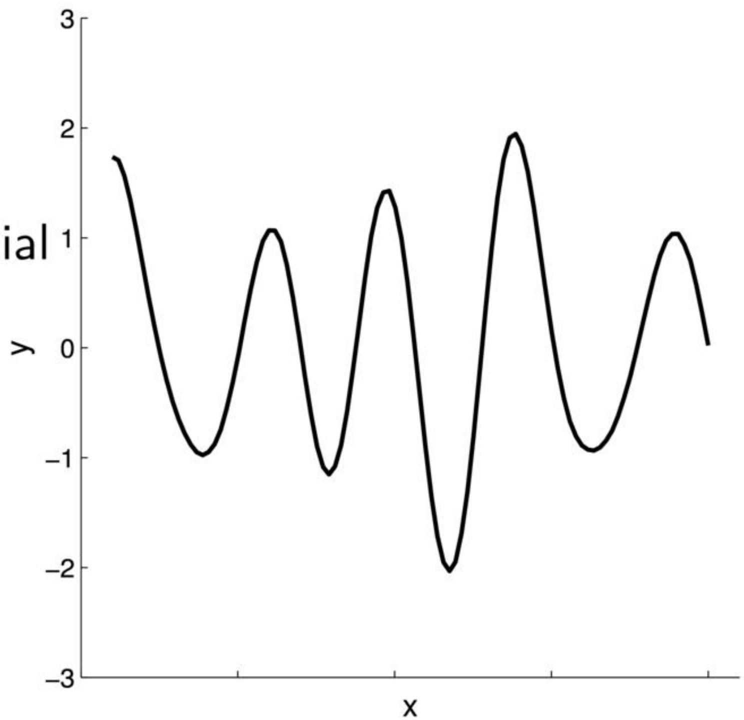


What effect do the Hyper-parameter have?

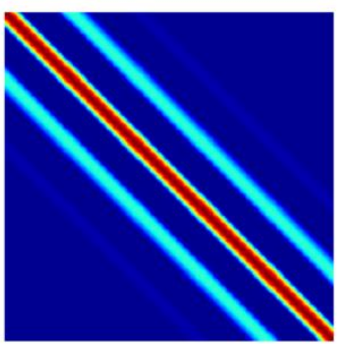
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



$\Sigma =$

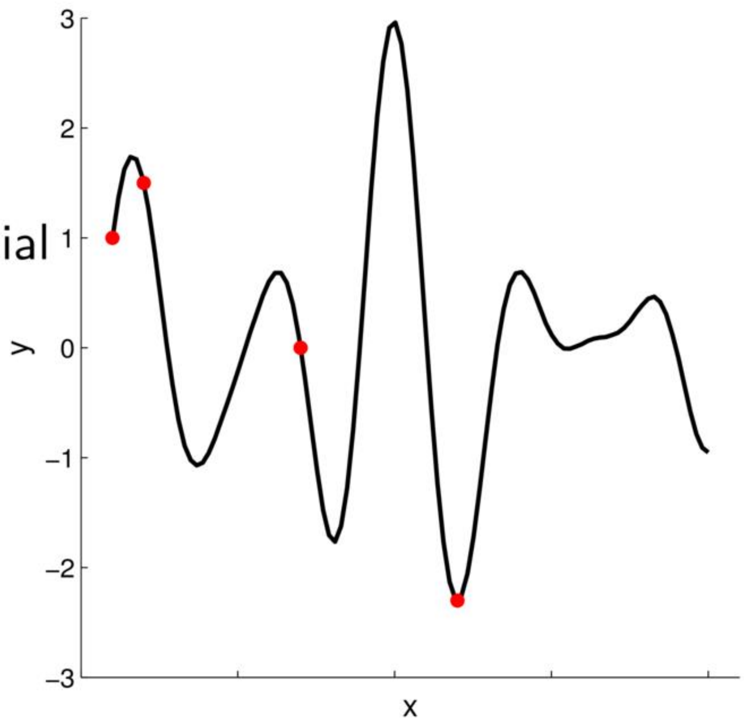


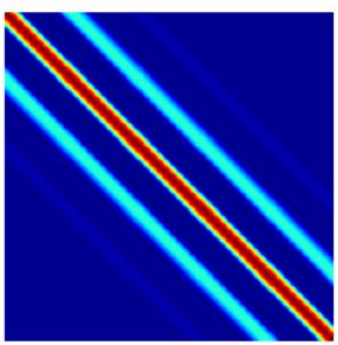
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



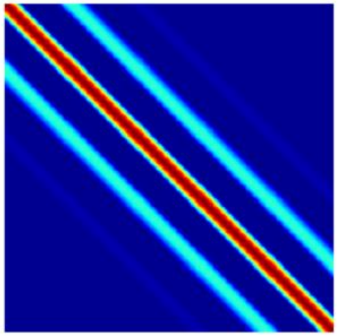
$\Sigma =$ 

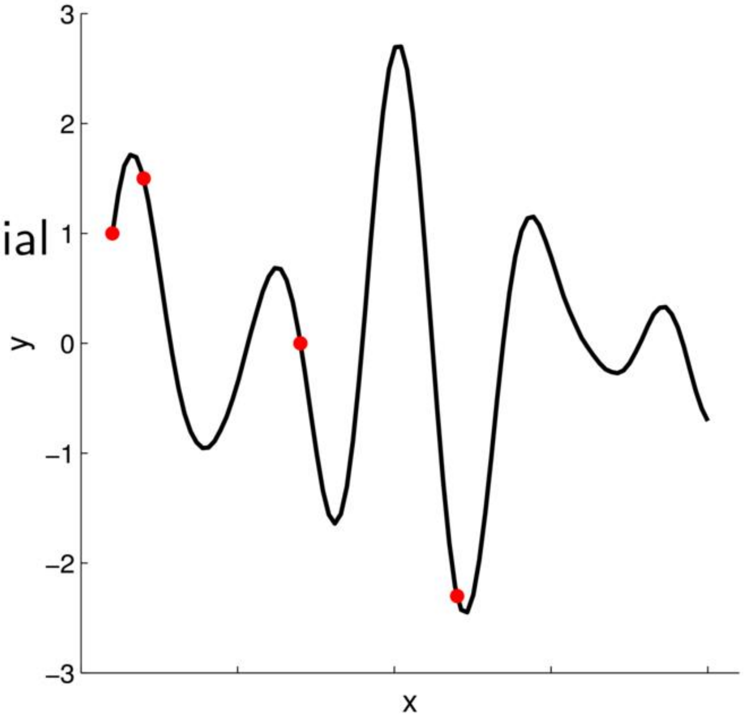
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 



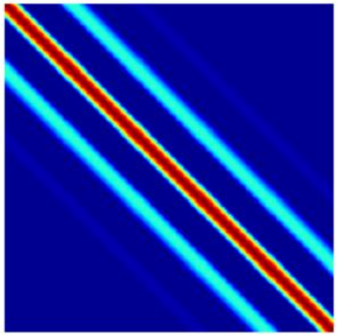
What effect do the Hyper-parameter have?

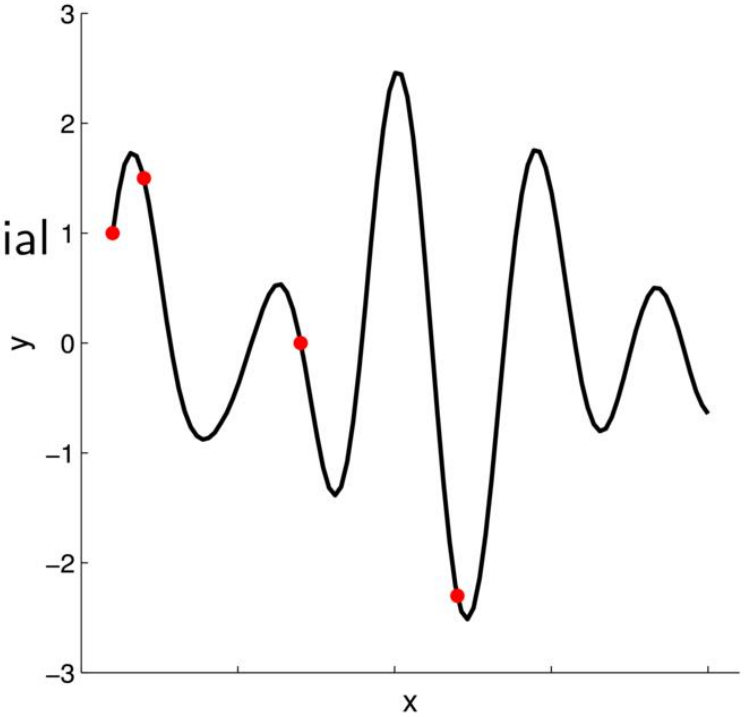
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$



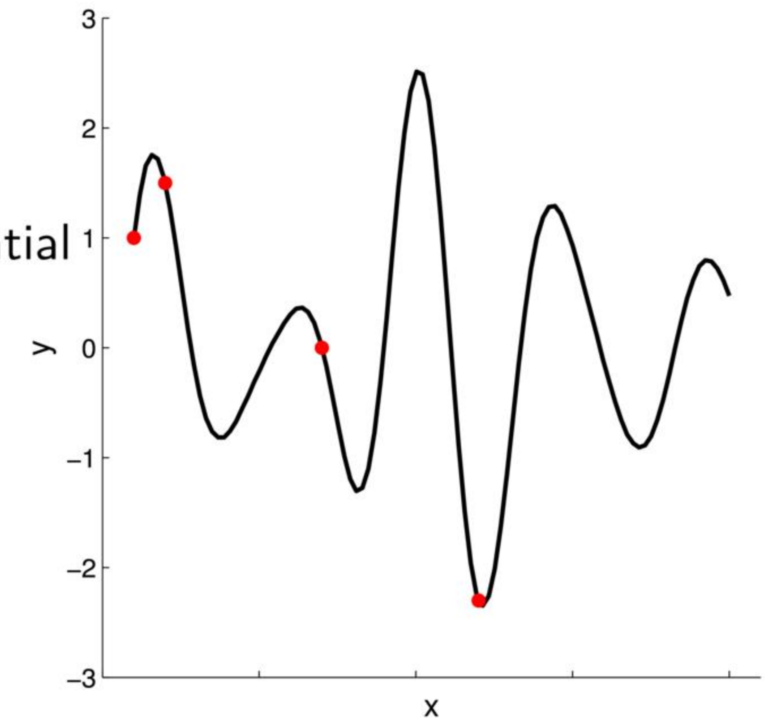


What effect do the Hyper-parameter have?

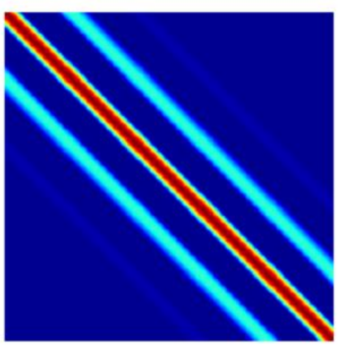
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



$\Sigma =$

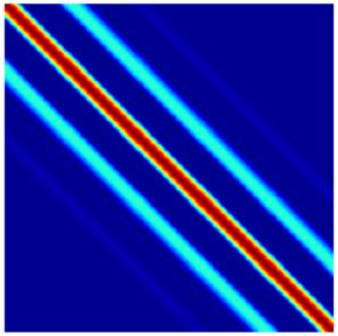


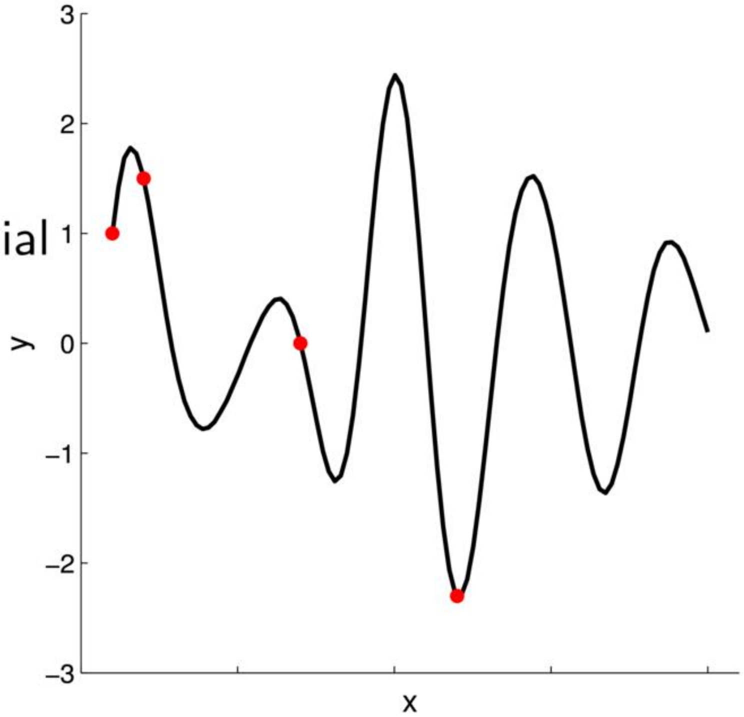
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

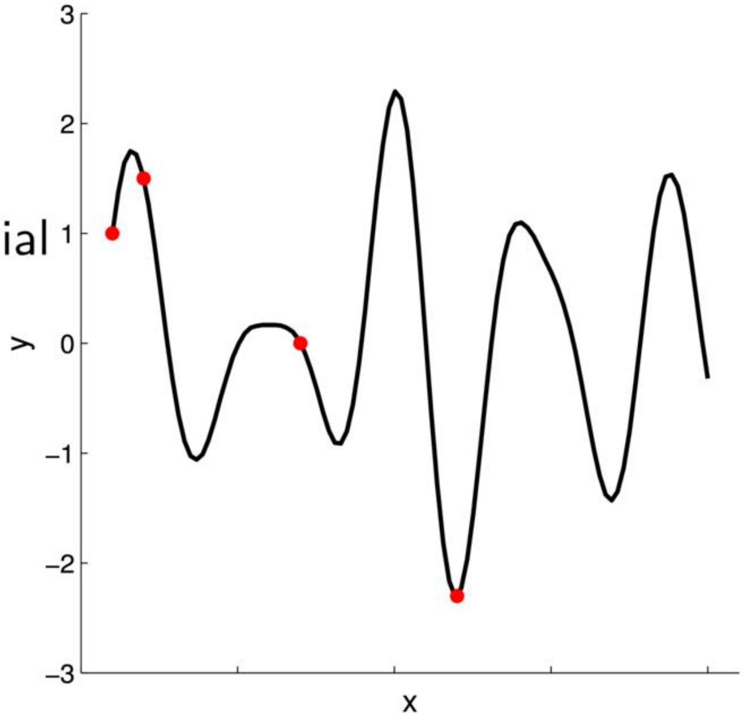


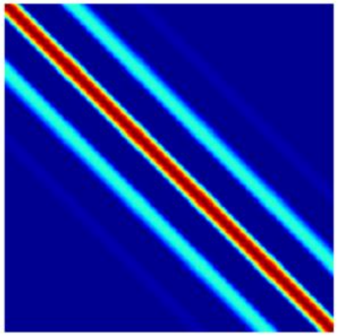
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



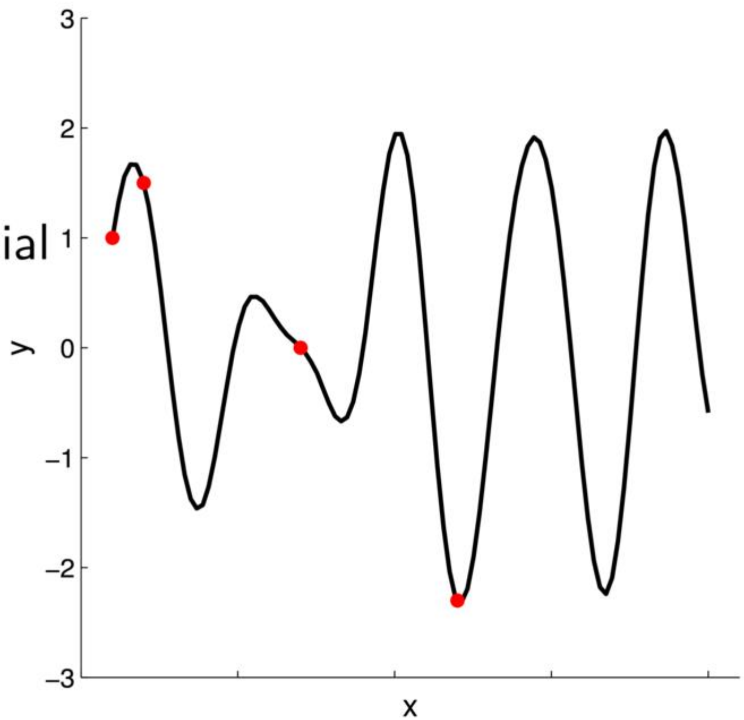
$\Sigma =$ 

What effect do the Hyper-parameter have?

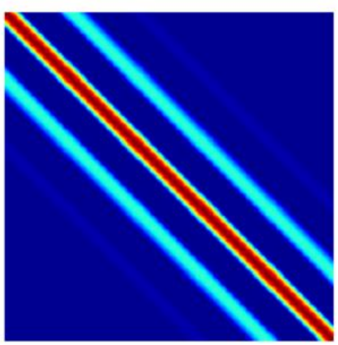
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



$\Sigma =$

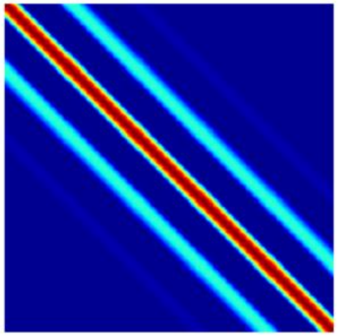


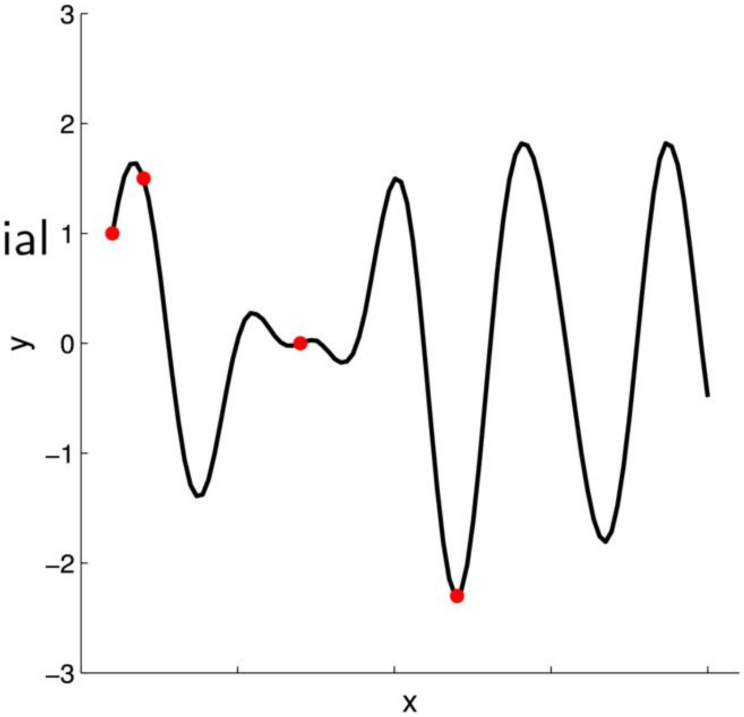
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

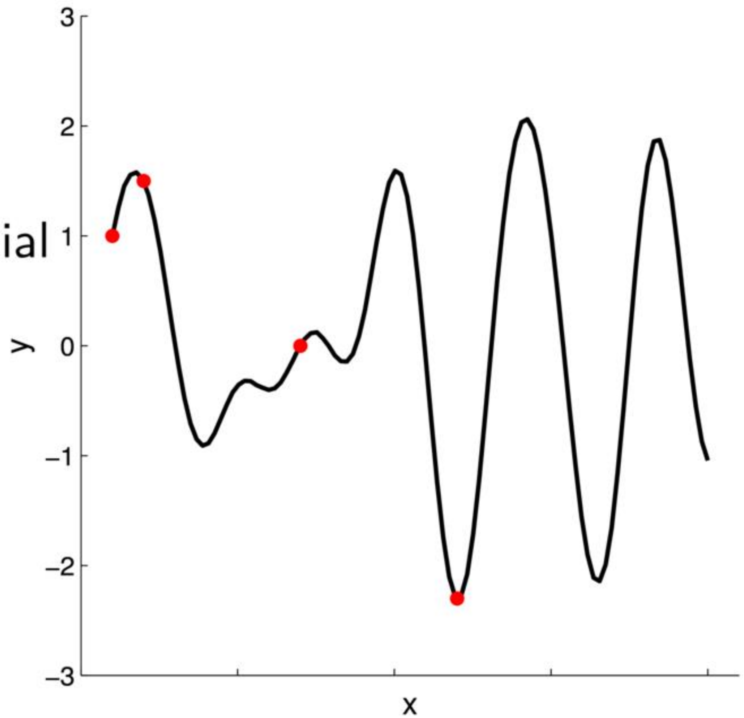


What effect do the Hyper-parameter have?

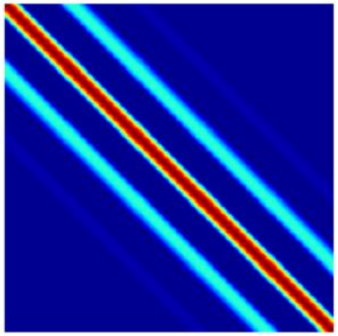
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



$\Sigma =$

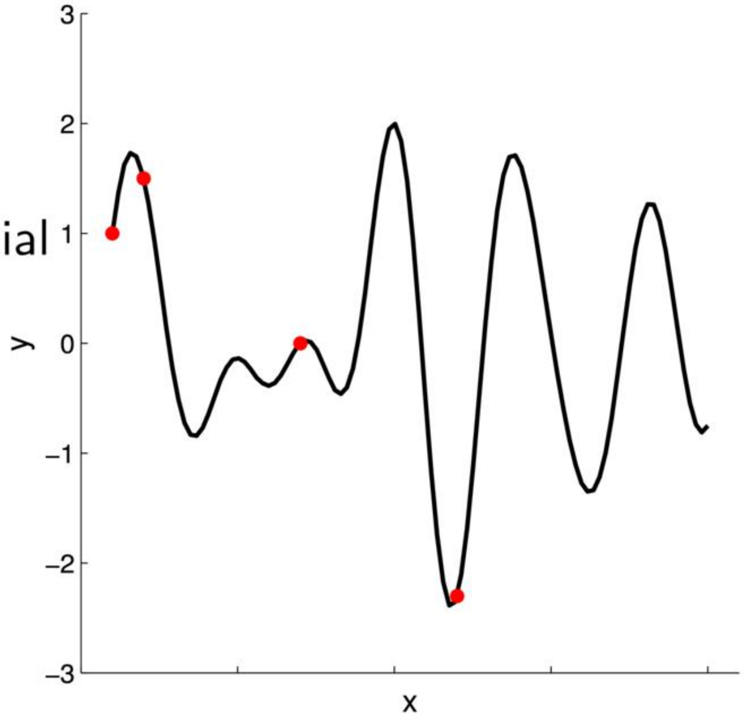


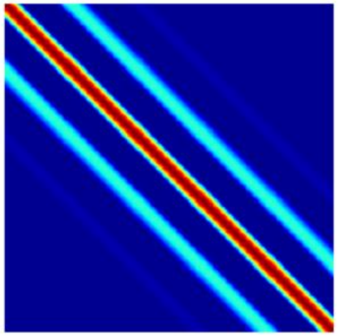
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential



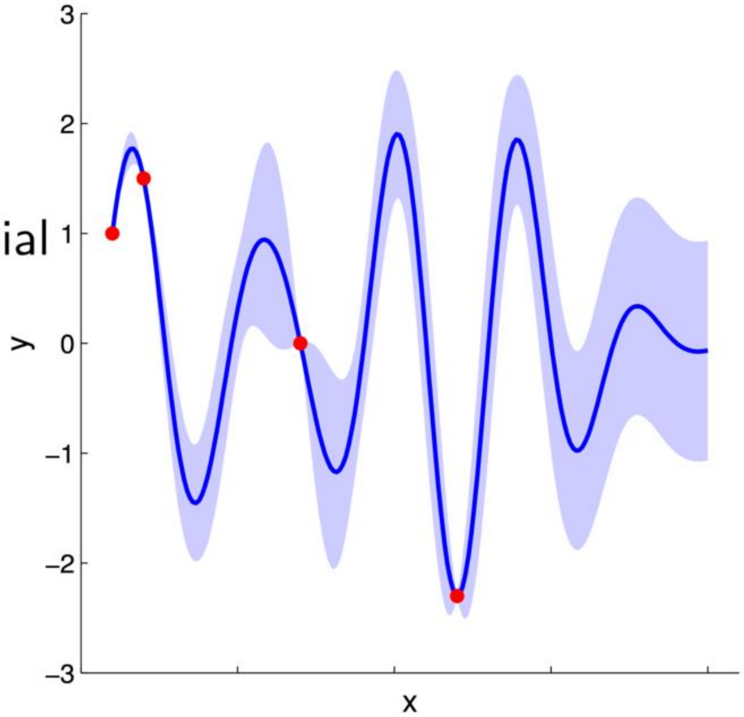
$\Sigma =$ 

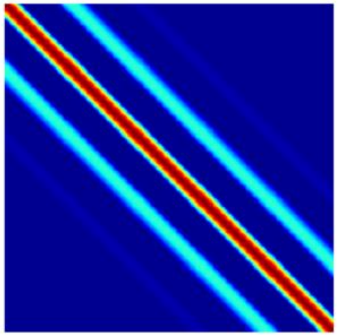
What effect do the Hyper-parameter have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

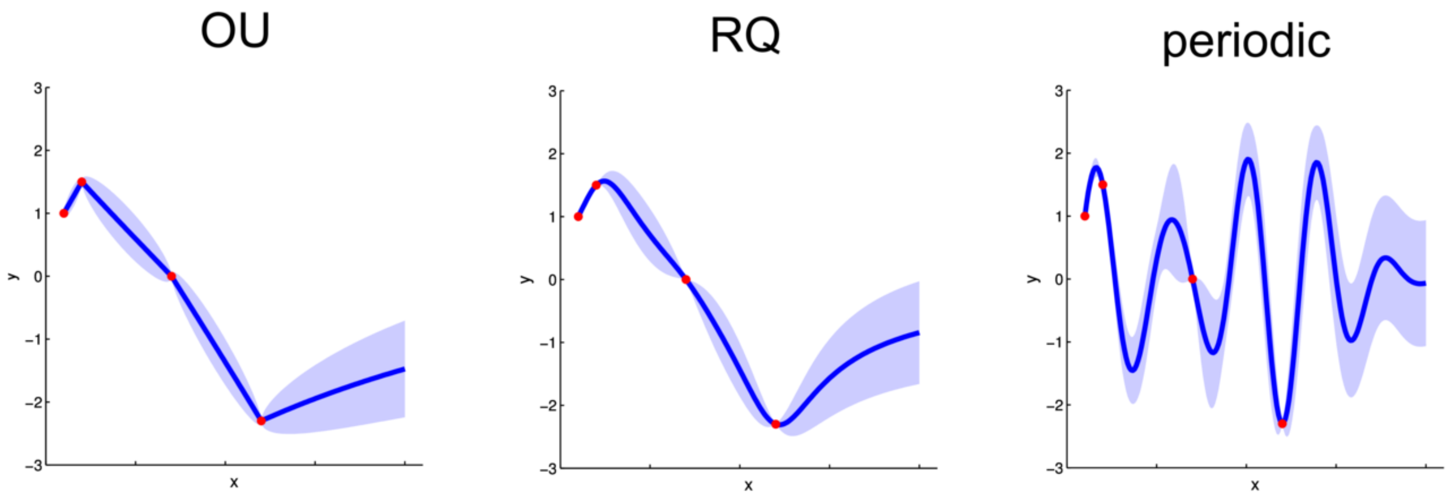
Periodic

sinusoid $\times$ squared exponential

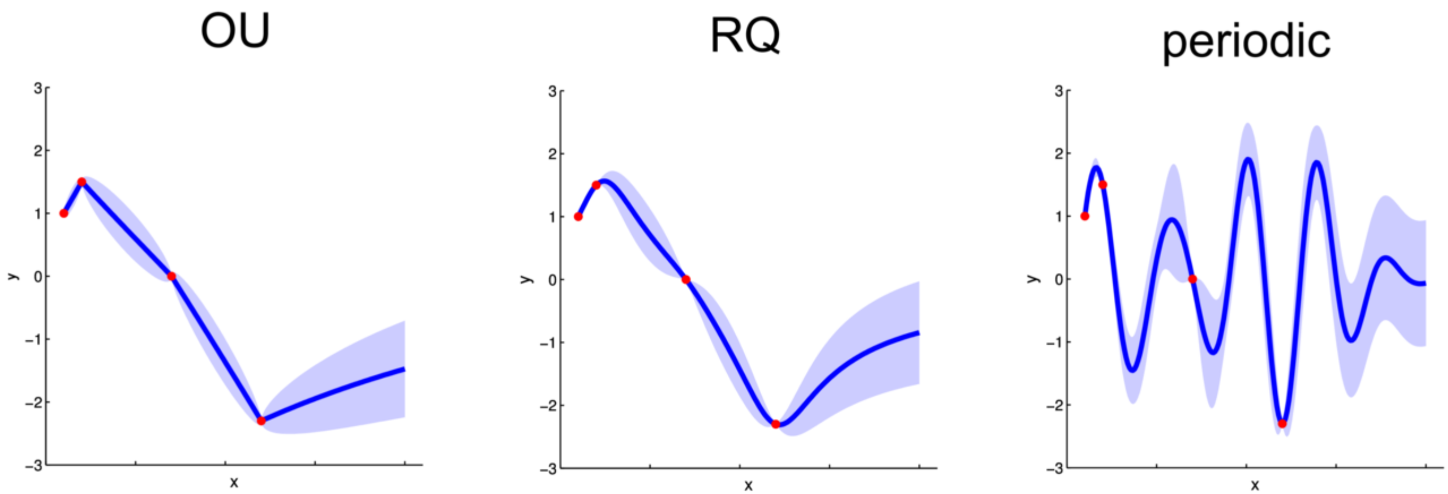


$\Sigma =$ 

The Covariance Function has a Large Effect



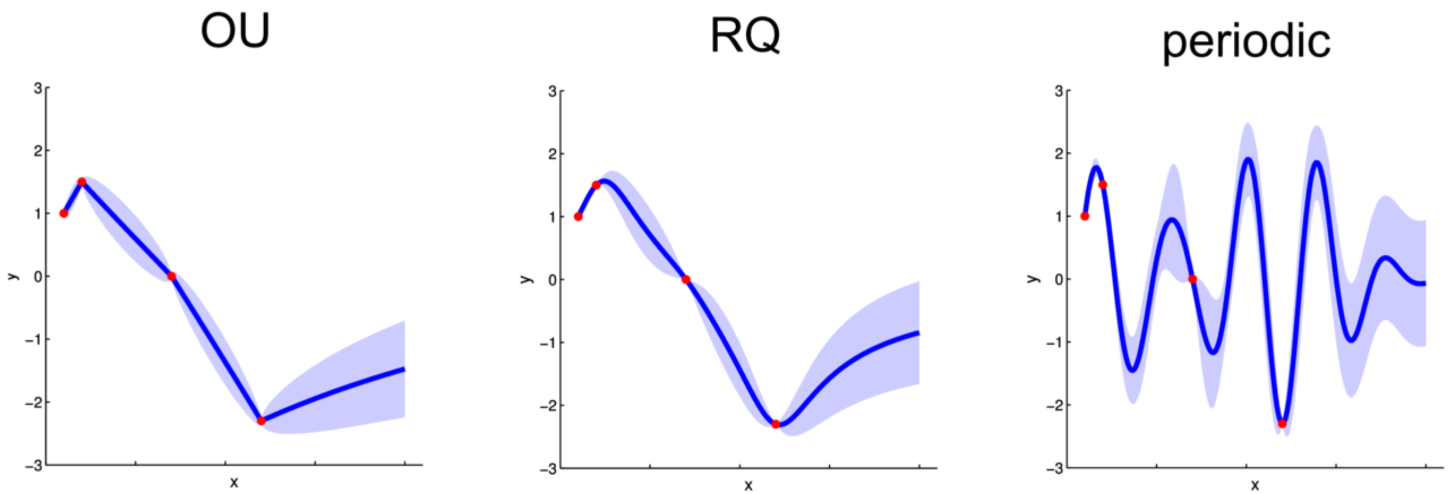
The Covariance Function has a Large Effect



Bayesian model comparison:

$$p(M|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|M)p(M)}{\sum_{M'} p(\mathbf{y}_{1:N}|M')p(M')}$$

The Covariance Function has a Large Effect

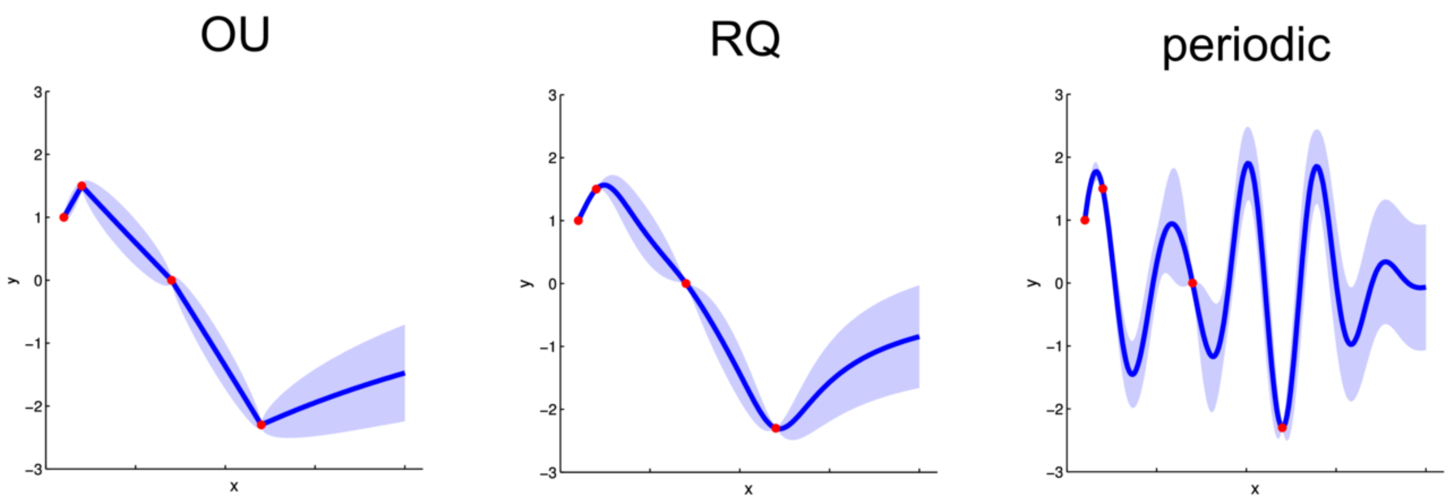


Bayesian model comparison:

$$p(M|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|M)p(M)}{\sum_{M'} p(\mathbf{y}_{1:N}|M')p(M')}$$

prior over models

The Covariance Function has a Large Effect



Bayesian model comparison:

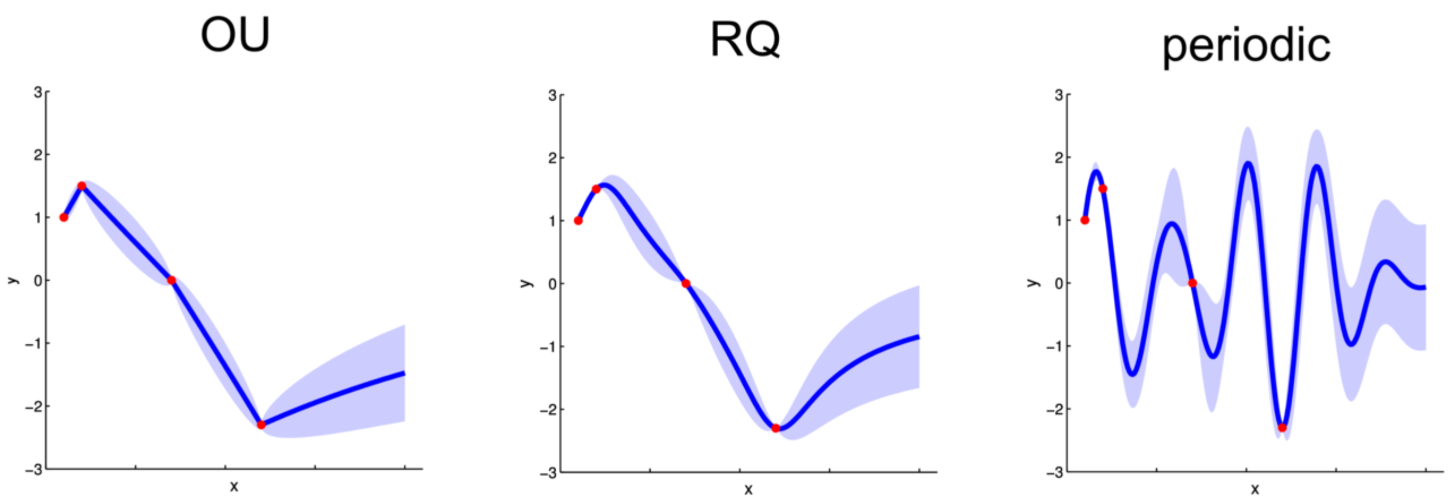
$$p(M|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|M)p(M)}{\sum_{M'} p(\mathbf{y}_{1:N}|M')p(M')}$$

prior over models

marginal likelihood

$$p(\mathbf{y}_{1:N}|M) = \int \mathrm{d}\theta \; p(\mathbf{y}_{1:N}|\theta, M)p(\theta|M)$$

The Covariance Function has a Large Effect



Bayesian model comparison:

$$p(M|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|M)p(M)}{\sum_{M'} p(\mathbf{y}_{1:N}|M')p(M')}$$

prior over models

marginal likelihood $p(\mathbf{y}_{1:N}|M) = \int \mathrm{d}\theta \; p(\mathbf{y}_{1:N}|\theta, M)p(\theta|M)$

Health warnings: Hard to compute (need approximations)
Often results are very sensitive to the priors $p(\theta|M)$