Homework 3

Regression, Gaussian Processes, and Boosting

Dana Van Aken

Problem 1: Gaussian Processes

(a) A comparison of covariance functions: see figures **??**, **??**, and **??**.

(b) Increasing σ^2 increases the "noisyness" of the output points, y_i. Figure **??** shows... TODO

(c) Show $p(x_1|x_2) \propto p(x_1, x_2)$

$$p(X_1|x_2) = \frac{1}{sqrt(2\pi)^k|\Sigma_{x_1|x_2}|} exp\Big(-\frac{1}{2}(x - \mu_{x_1|x_2})\Big)$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma 12 \\ \Sigma 21 & \Sigma 22 \end{bmatrix}\right)$$

Let $\Sigma^{-1} = \Lambda^{-1}$ such that

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} = \Lambda$$

We can focus on the exponent since we want to find $\mu_{x_1|x_2}$ and $\Sigma_{x_1|x_2}$.

$$\begin{aligned} exp &= -\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) \\ &= -\frac{1}{2}\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \\ &= -\frac{1}{2}\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^T \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \end{aligned}$$

$$= -\frac{1}{2}(x_1 - \mu_1)^T \Lambda_{11}(x_1 - \mu_1) - \frac{1}{2}(x_1 - \mu_1)^T \Lambda_{12}(x_2 - \mu_2)$$
$$- \frac{1}{2}(x_2 - \mu_2)^T \Lambda_{21}(x_1 - \mu_1) - \frac{1}{2}(x_2 - \mu_2)^T \Lambda_{22}(x_2 - \mu_2)$$

We can call the last term, $-\frac{1}{2}(x_2 - \mu_2)^T \Lambda_{22}(x_2 - \mu_2)$, C since it does not depend on x_1 (constant).

$$= -\frac{1}{2}(x_1 - \mu_1)^T \Lambda_{11}(x_1 - \mu_1) - \frac{1}{2}(x_1 - \mu_1)^T \Lambda_{12}(x_2 - \mu_2) - \frac{1}{2}(x_2 - \mu_2)^T \Lambda_{21}(x_1 - \mu_1) + C$$

$$= -\frac{1}{2}x_1^T \Lambda_{11} x_1 + \frac{1}{2}x_1^T \Lambda_{11} \mu_1 + \frac{1}{2}\mu_1^T \Lambda_{11} x_1 - \frac{1}{2}\mu_1^T \Lambda_{11} \mu_1$$
$$- \frac{1}{2}x_1^T \Lambda_{12} x_2 + \frac{1}{2}x_1^T \Lambda_{12} \mu_2 + \frac{1}{2}\mu_1^T \Lambda_{12} x_2 - \frac{1}{2}\mu_1^T \Lambda_{12} \mu_2$$
$$- \frac{1}{2}x_2^T \Lambda_{21} x_1 + \frac{1}{2}x_2^T \Lambda_{21} \mu_1 + \frac{1}{2}\mu_2^T \Lambda_{21} x_1 - \frac{1}{2}\mu_2^T \Lambda_{21} \mu_1 + C$$

Again, include any constants that do not depend on x_1 in C.

$$= -\frac{1}{2}x_1^T \Lambda_{11} x_1 + \frac{1}{2}x_1^T \Lambda_{11} \mu_1 + \frac{1}{2}\mu_1^T \Lambda_{11} x_1 - \frac{1}{2}x_1^T \Lambda_{12} x_2 + \frac{1}{2}x_1^T \Lambda_{12} \mu_2 - \frac{1}{2}x_2^T \Lambda_{21} x_1 + \frac{1}{2}\mu_2^T \Lambda_{21} x_1 + C$$

We can use the fact that $\Lambda_{21} = \Lambda_{12}^T$ to reduce the equation.

$$= -\frac{1}{2}x_1^T \Lambda_{11} x_1 + x_1^T \Lambda_{11} \mu_1 - x_1^T \Lambda_{12} x_2 + x_1^T \Lambda_{12} \mu_2 + C$$

(d)

(e) A comparison of covariance functions sampled from $p(f(X)|Y_*)$: see figures **??**, **??**, and **??**.

(f) TODO

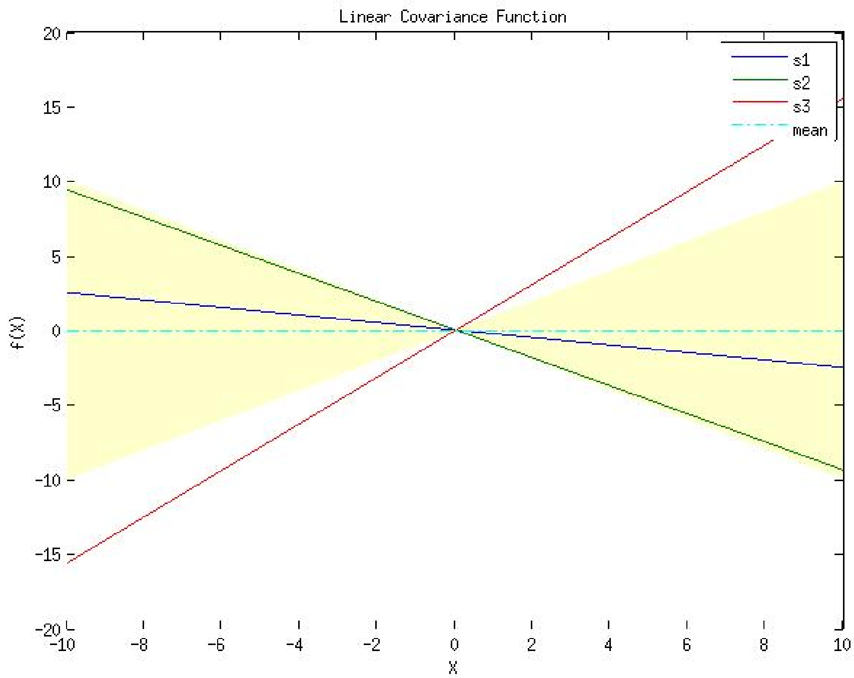

Figure 1: Linear Covariance Function

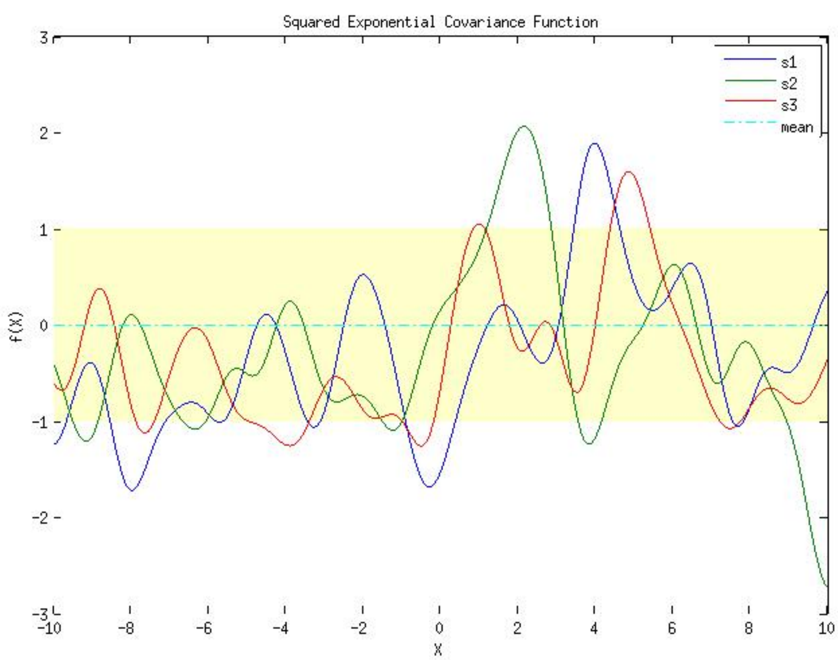

Figure 2: Square Exponential Covariance Function

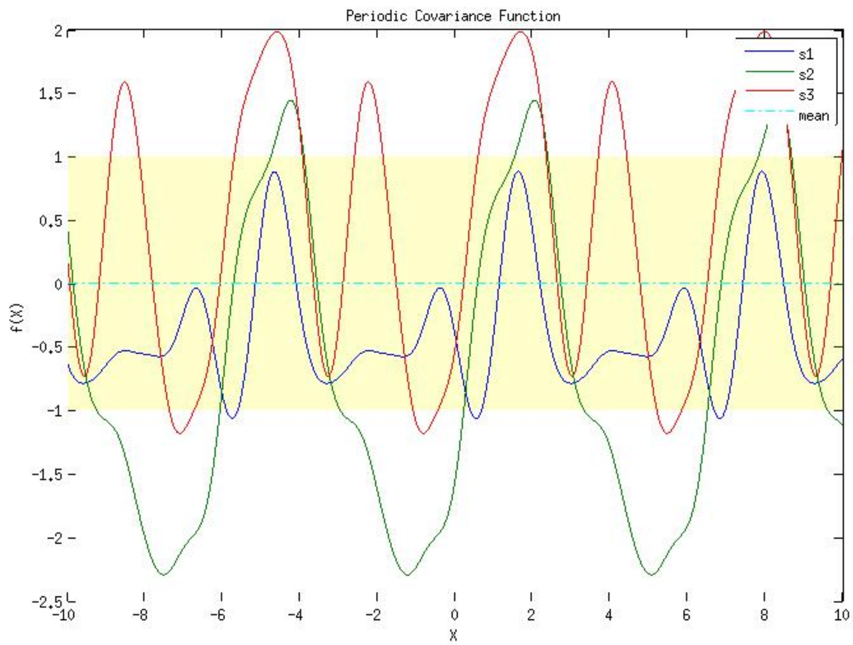

Figure 3: Periodic Covariance Function

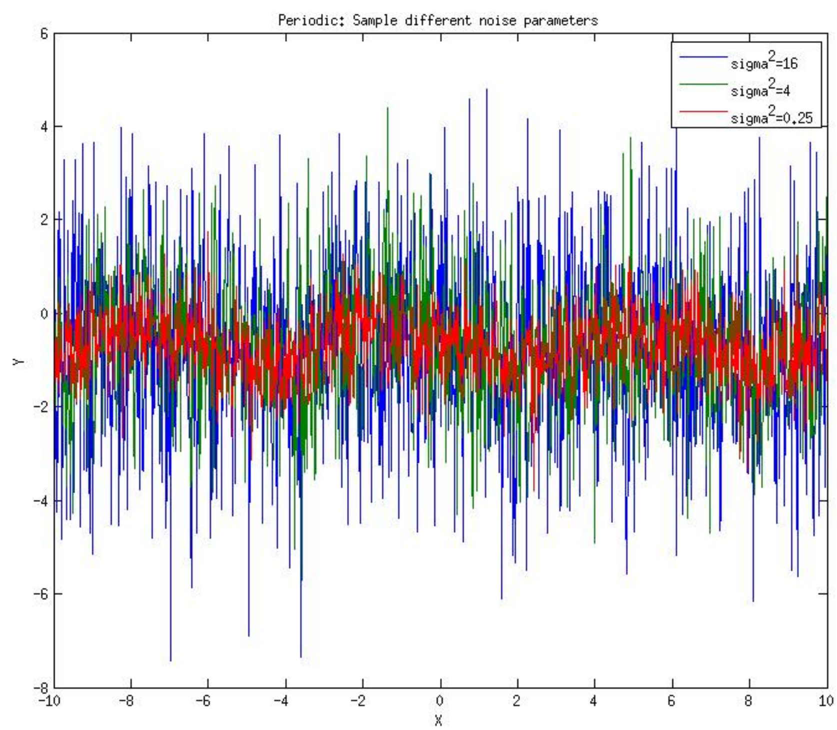

Figure 4: Sampling Different Gaussian Noise Parameters Using a Periodic Covariance Function

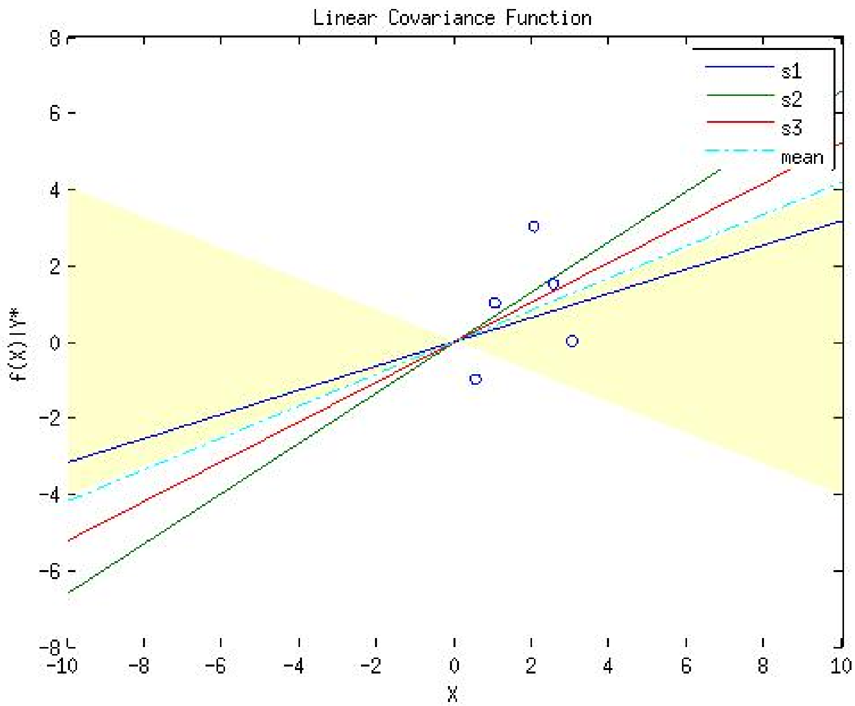

Figure 5: Linear Covariance Function

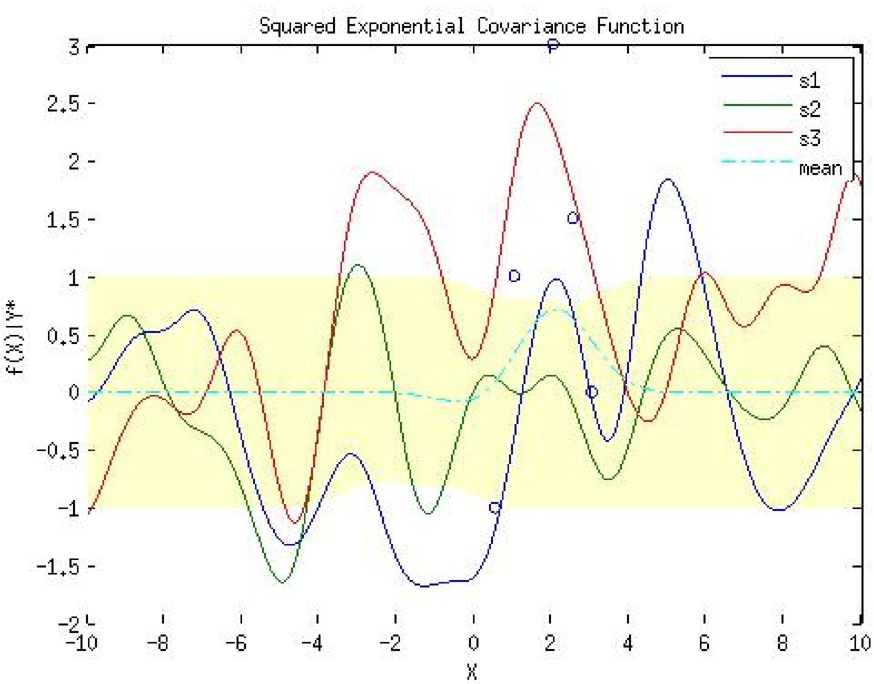

Figure 6: Square Exponential Covariance Function

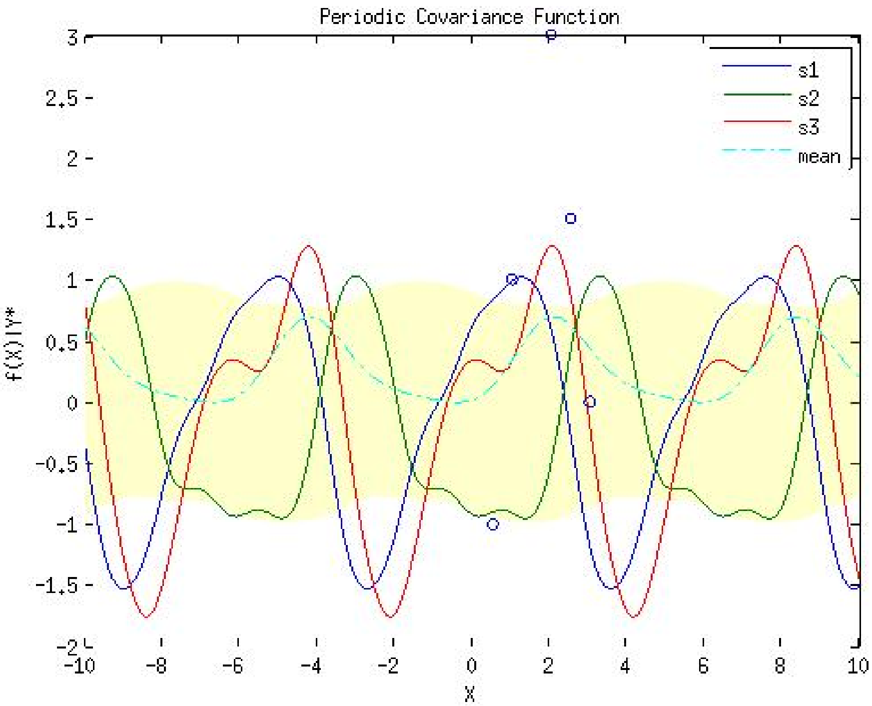

Figure 7: Periodic Covariance Function

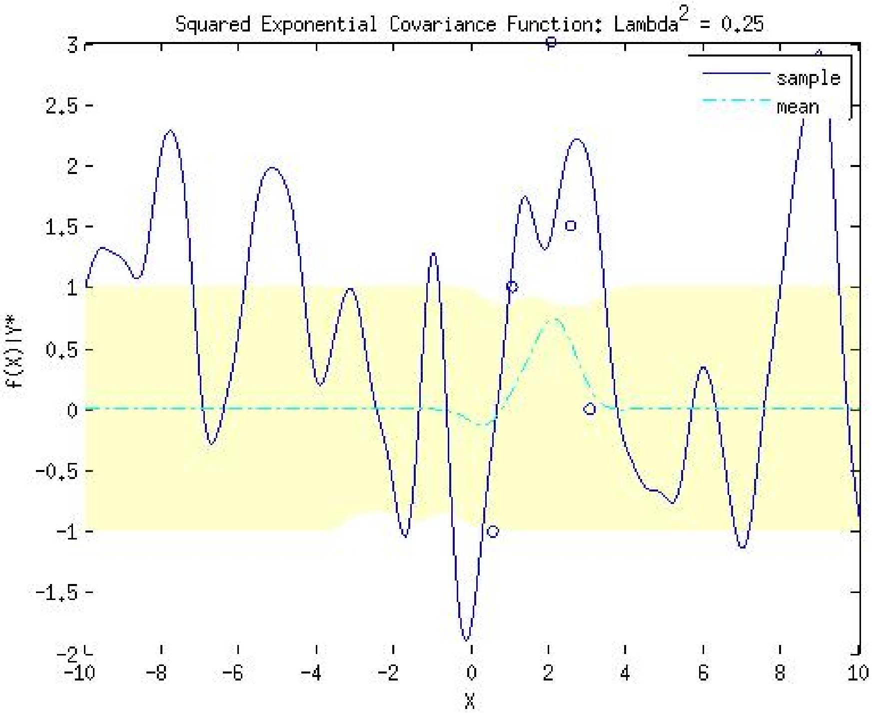

Figure 8: Sampling Different λ^2 Parameters Using the Squared Exponential Function

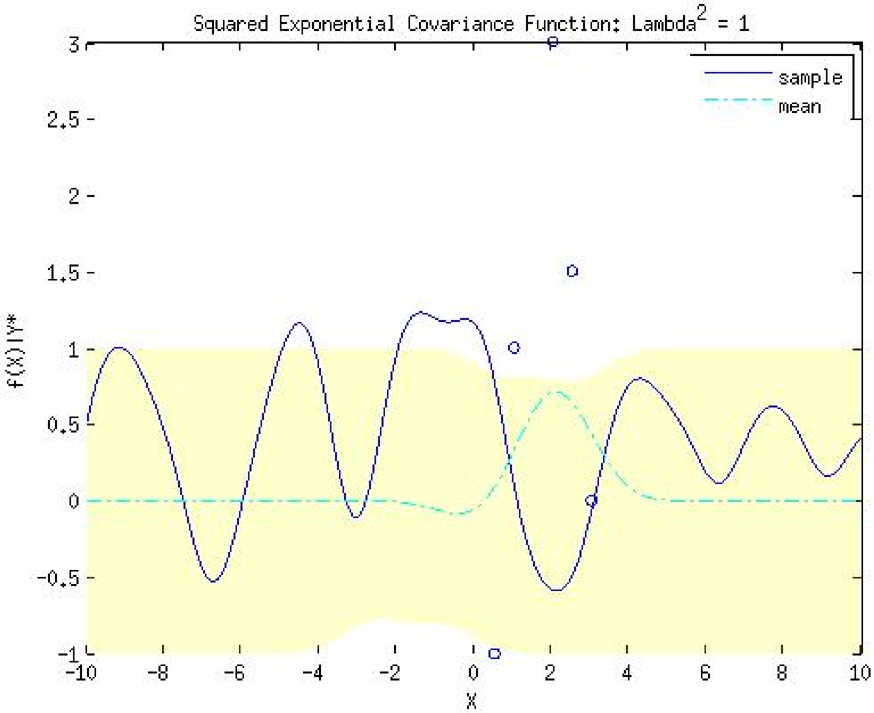

Figure 9: Sampling Different λ^2 Parameters Using the Squared Exponential Function

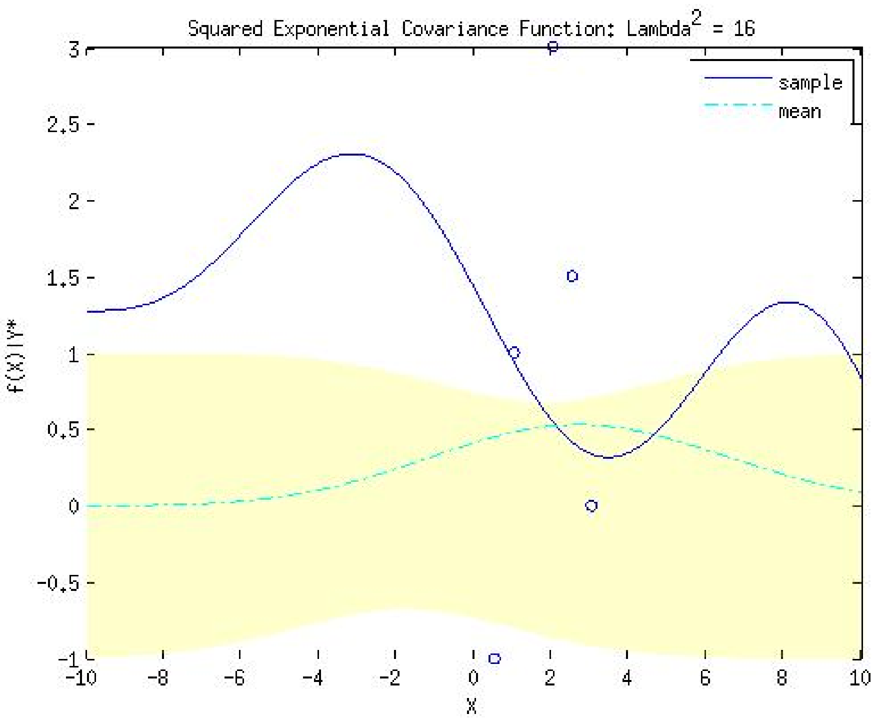

Figure 10: Sampling Different λ^2 Parameters Using the Squared Exponential Function