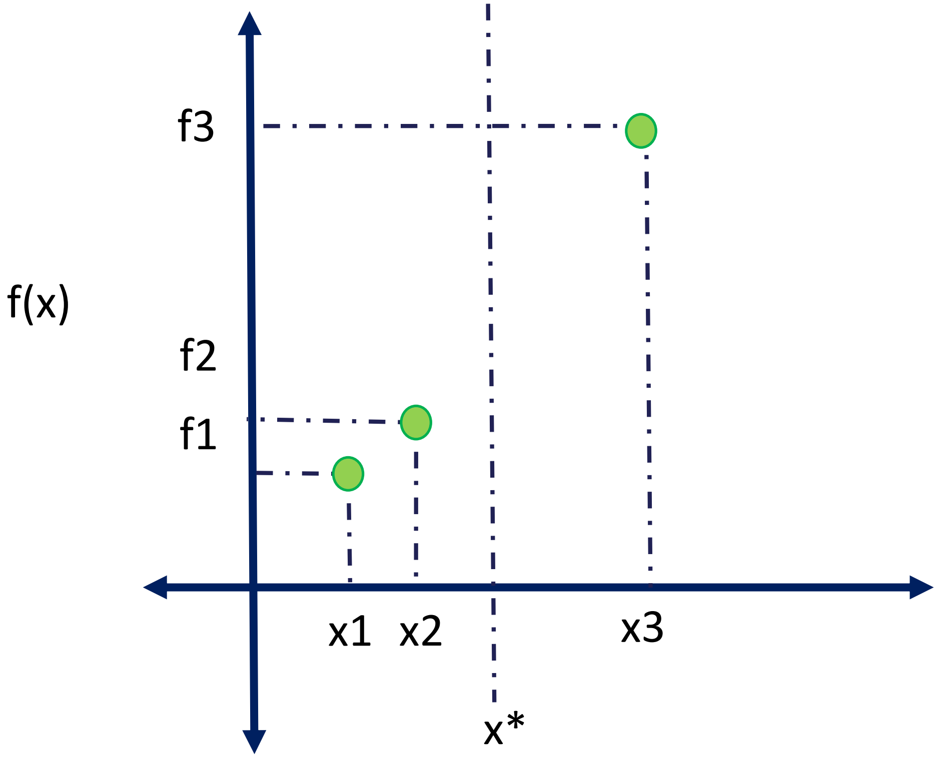
Train In Data

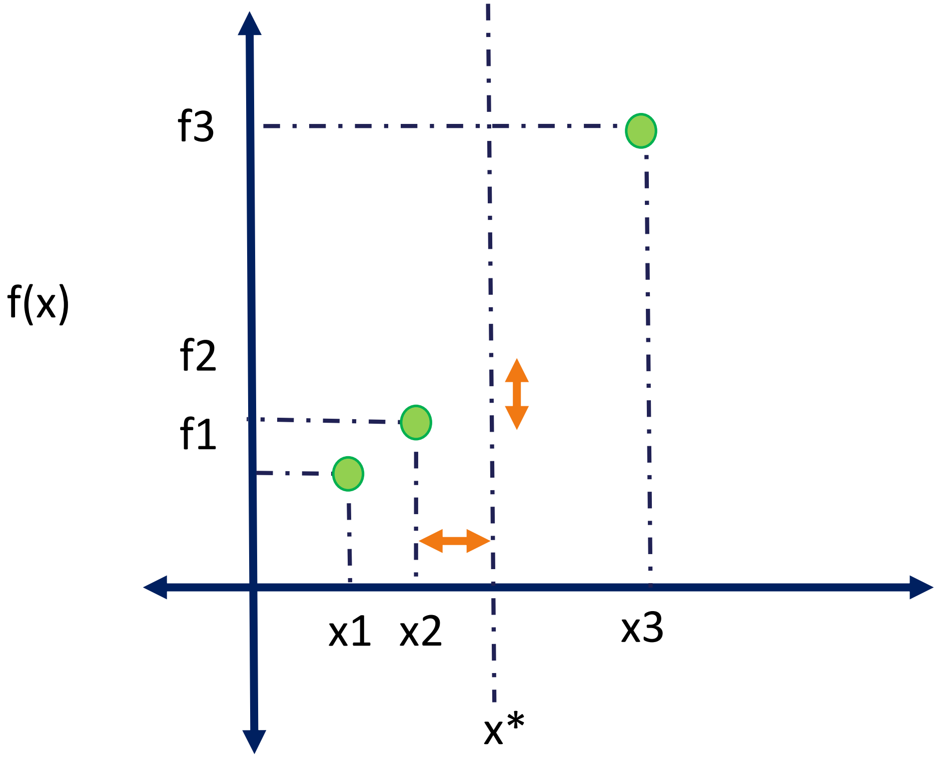
Kernels

Similarity



$$\begin{bmatrix} f1 \\ f2 \\ f3 \end{bmatrix} \sim N(\mu = \begin{bmatrix} m1 \\ m2 \\ m3 \end{bmatrix}, \Sigma = \begin{bmatrix} K11 & K12 & K13 \\ K21 & K22 & K23 \\ K31 & K32 & K33 \end{bmatrix})$$

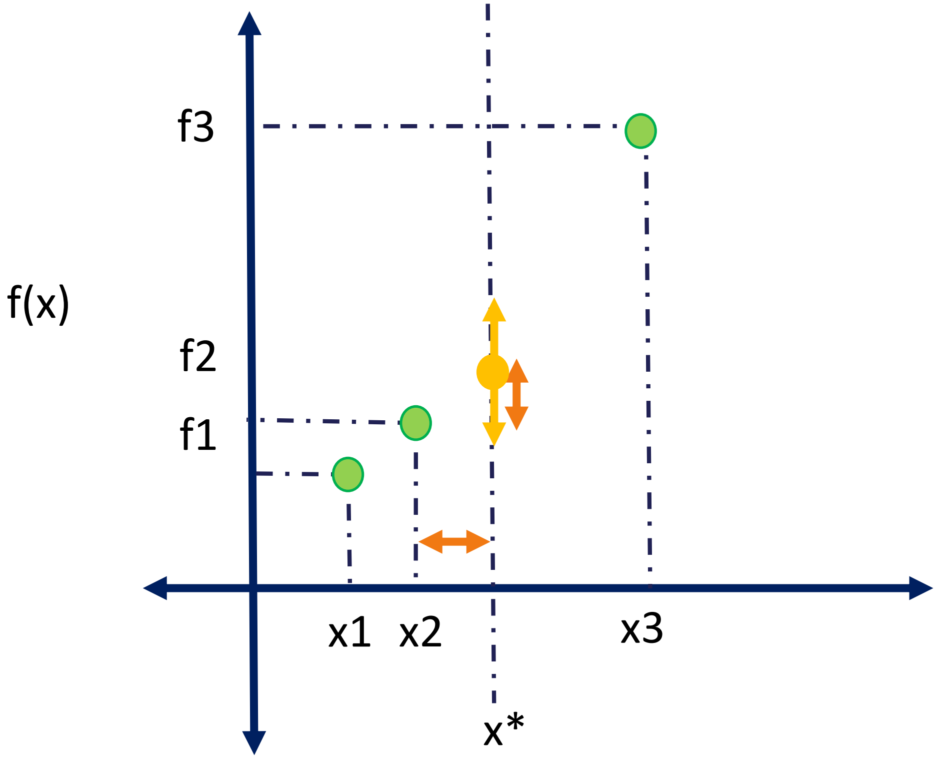
Similarity



$$\begin{bmatrix} f1 \\ f2 \\ f3 \end{bmatrix} \sim N(\mu = \begin{bmatrix} m1 \\ m2 \\ m3 \end{bmatrix}, \Sigma = \begin{bmatrix} K11 & K12 & K13 \\ K21 & K22 & K23 \\ K31 & K32 & K33 \end{bmatrix})$$

If the distance in ***x*** is small, we expect the distance in ***y*** to be small.

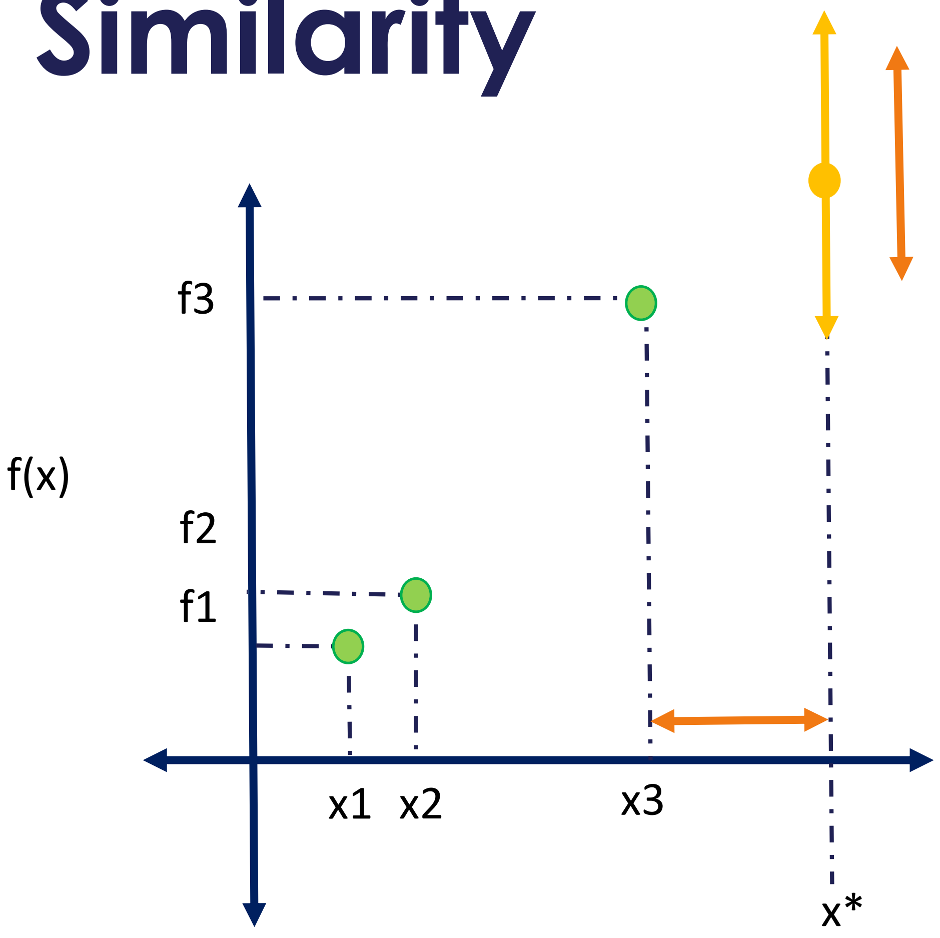
Similarity



$$\begin{bmatrix} f1 \\ f2 \\ f3 \end{bmatrix} \sim N(\mu = \begin{bmatrix} m1 \\ m2 \\ m3 \end{bmatrix}, \Sigma = \begin{bmatrix} K11 & K12 & K13 \\ K21 & K22 & K23 \\ K31 & K32 & K33 \end{bmatrix})$$

If the distance in ***x*** is small, we expect the distance in ***y*** to be small.

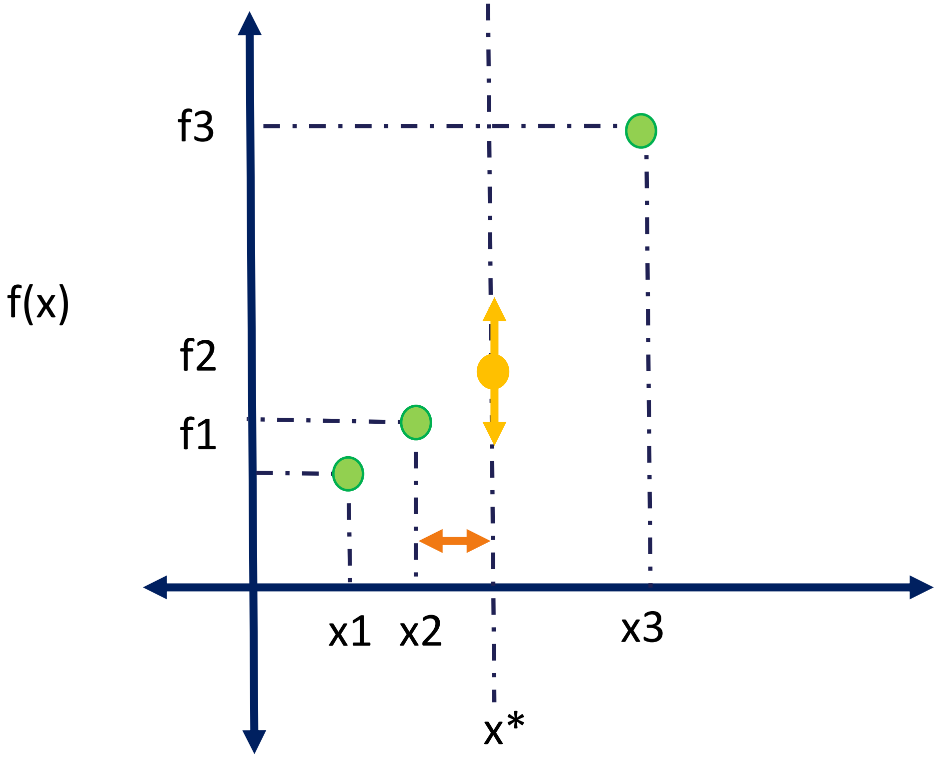
Similarity



$$\begin{bmatrix} f1 \\ f2 \\ f3 \end{bmatrix} \sim N(\mu = \begin{bmatrix} m1 \\ m2 \\ m3 \end{bmatrix}, \Sigma = \begin{bmatrix} K11 & K12 & K13 \\ K21 & K22 & K23 \\ K31 & K32 & K33 \end{bmatrix})$$

If the distance in ***x*** is big, we expect the distance in ***y*** to be big.

Kernels

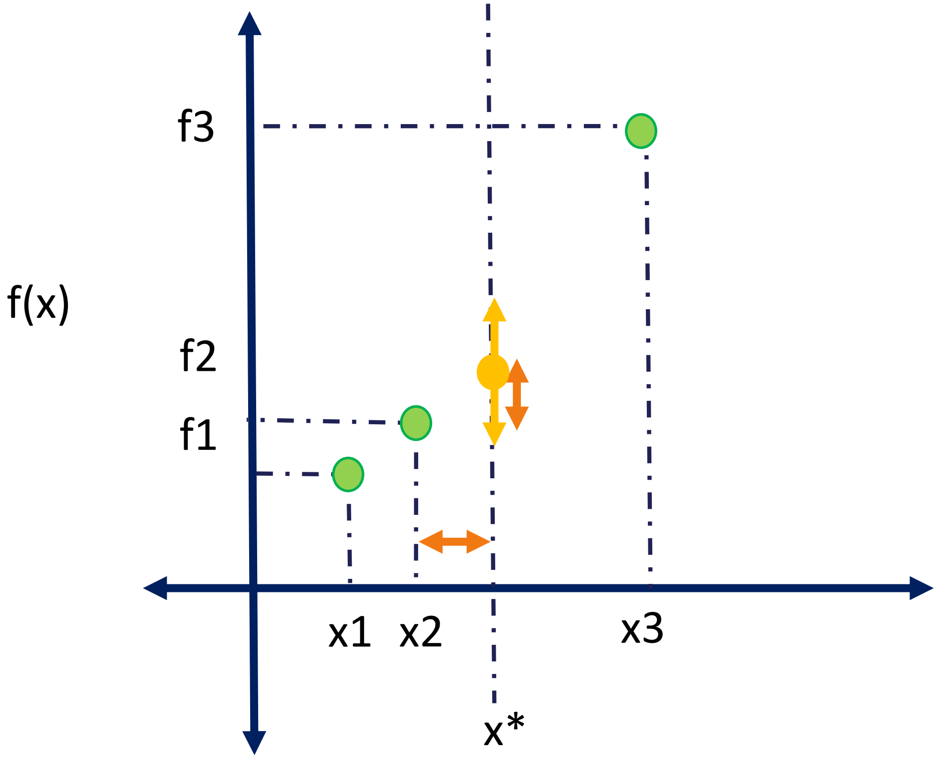


$$\begin{bmatrix} f1 \\ f2 \\ f3 \end{bmatrix} \sim N(\mu = \begin{bmatrix} m1 \\ m2 \\ m3 \end{bmatrix}, \Sigma \begin{bmatrix} K11 & K12 & K13 \\ K21 & K22 & K23 \\ K31 & K32 & K33 \end{bmatrix})$$

Kernel:

- K(xi, xj)
- Kernels property ➔ points closer in the input space are more strongly correlated.
- Crucial to determine smoothness of the GP

Squared exponential kernel



$$\begin{bmatrix} f1 \\ f2 \\ f3 \end{bmatrix} \sim N(\mu = \begin{bmatrix} m1 \\ m2 \\ m3 \end{bmatrix}, \Sigma = \begin{bmatrix} K11 & K12 & K13 \\ K21 & K22 & K23 \\ K31 & K32 & K33 \end{bmatrix})$$

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{1}{2}\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2\right)$$

- If x_i and x_j are close, x_i-x_j ~ 0 ➔ $k(x_i, x_j)$ =1
- If x_i and x_j are different, x_i-x_j is big ➔ $k(x_i, x_j)$ = 0

Squared exponential kernel

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{1}{2\theta^2}\|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$$

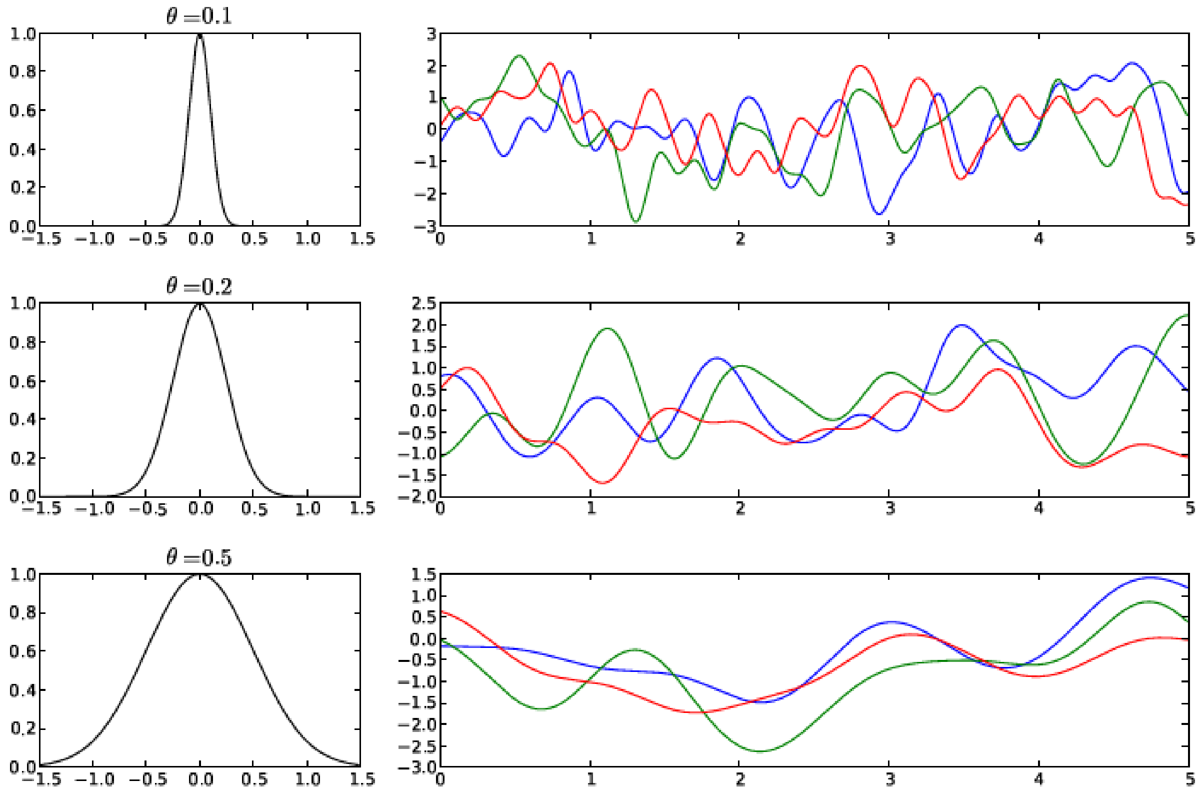


Figure 3: *The effect of changing the kernel hyperparameters. Shown are squared exponential kernels with* $\theta = 0.1, 0.2, 0.5$. *On the left is the function* $k(0, \mathbf{x})$. *On the right are some one-dimensional functions sampled from a GP with the hyperparameter value.*

Image taken from Bochu, Cora, de Freitas, 2010

Squared exponential kernel

$$k(\mathbf{x}_i, \mathbf{x}_j) = \alpha \exp\left(-\frac{1}{2\theta^2} \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$$

- Kernel hyperparameters can be determined with cross-validation or Bayesian optimization

Martérn kernel

$$k(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{2^{\varsigma-1}\Gamma(\varsigma)} \left(2\sqrt{\varsigma}\|\mathbf{x}_i - \mathbf{x}_j\|\right)^{\varsigma} H_{\varsigma}\left(2\sqrt{\varsigma}\|\mathbf{x}_i - \mathbf{x}_j\|\right),$$

where $\Gamma(\cdot)$ and $H_{\varsigma}(\cdot)$ are the Gamma function and the Bessel function of order ς.

- ζ is a smoothness parameter that allows greater flexibility to model the GP

THANK YOU

www.trainindata.com