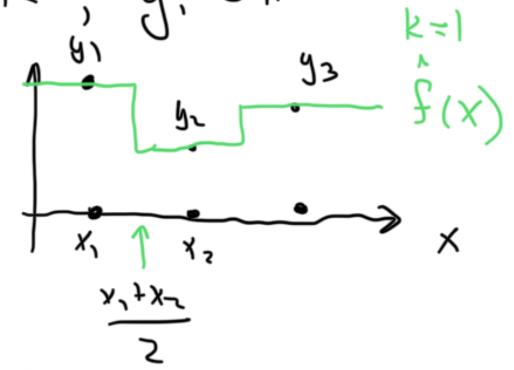
Gaussian process regression

$$D = \{(\vec{x}_i, y_i)\}_{i=1}^{m}, \quad \vec{x}_i \in \mathbb{R}^d, \quad y_i \in \mathbb{R}$$

$$\hat{f}(\vec{x}) = \frac{1}{k} \sum_{i \in NN_k(\vec{x})} f(\vec{x}_i)$$



Kernel regression:

1. $d_i = \|\vec{x} - \vec{x}_i\|_2^2$. distance

2. $s_i = \exp(-d_i)$. similarity $\quad \exp\left(-\frac{\|\vec{x}-\vec{x}_i\|^2}{1}\right) = k(\vec{x}, \vec{x}_i)$

3. $w_i = \frac{s_i}{\sum_{j=1}^{m} s_j}$

$$f(\vec{x}) = \sum_i w_i f(\vec{x}_i) = \sum_i w_i y_i$$

NN regression $\rightarrow \sum_i w_i y_i$, $\quad w_i = \begin{cases} 1/k, & \vec{x}_i \in kNN(\vec{x}) \\ 0, & \text{otherwise} \end{cases}$

$$s_i = \begin{cases} 1, & \vec{x}_i \in kNN(\vec{x}) \\ 0, & \text{otherwise} \end{cases}$$

$$w_i = \frac{s_i}{\sum s_i} = \frac{s_i}{k}$$

more clever!

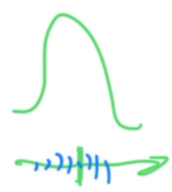
$$\hat{f}(\vec{x}) = \sum w_i y_i$$

Gaussian random variable

$$y \sim N(0, \sigma^2)$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right)$$

$$\Sigma = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{pmatrix}, \quad \sigma_{12}^2 = \sigma_{21}^2$$

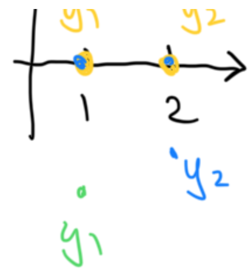


$$\sigma_{ij}^2 = \mathbb{E}\, y_i y_j$$

$$p(y_1 \mid y_2) = N(\mu_1, \sigma_1^2)$$

$$\mu_1 = \sigma_{12}^2 \frac{1}{\sigma^2} y_2$$

$$\sigma_1^2 = \sigma_{11}^2 - \frac{\sigma_{12}^4}{\sigma_{22}^2} \geq 0 \qquad K_{m+1}$$



$$\begin{array}{c} \\ \begin{pmatrix} y_1 \\ \vdots \\ y_m \\ y \end{pmatrix} \end{array} = \begin{pmatrix} \vec{y} \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} \vec{0} \\ 0 \end{pmatrix}, \begin{array}{cc} m & 1 \\ \multicolumn{2}{c}{\begin{pmatrix} K & \vec{k} \\ \vec{k}^T & k \end{pmatrix}} \end{array} \right) \quad V$$

$\mathbb{E}\, y^2$

$$p(y \mid \vec{y}) = N\Big(\underbrace{\underset{1\times m}{\vec{k}^T} \underset{m\times m}{K^{-1}} \vec{y}}_{}, \ \underbrace{k - \vec{k}^T K^{-1} \vec{k}}_{} \Big)$$

$$\sum_{i=1}^{m} \alpha_i y_i$$

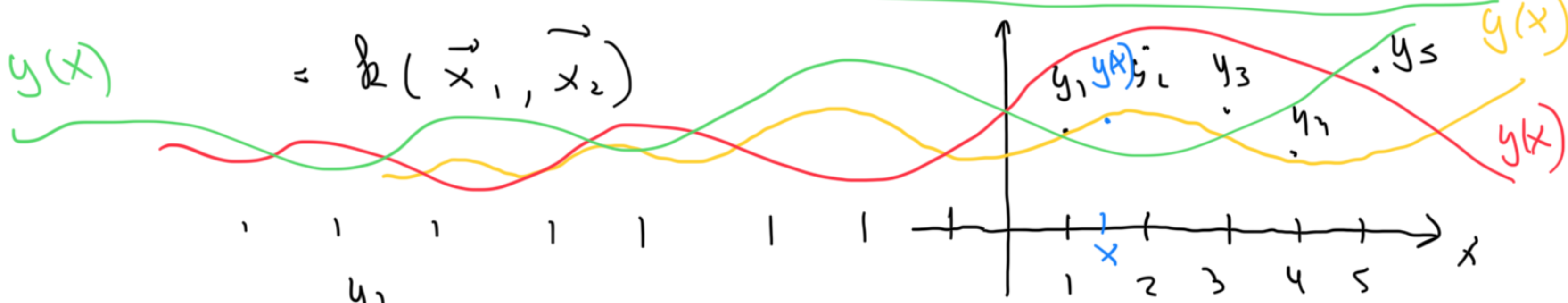
$$\vec{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix} = \vec{k}^T K^{-1}$$

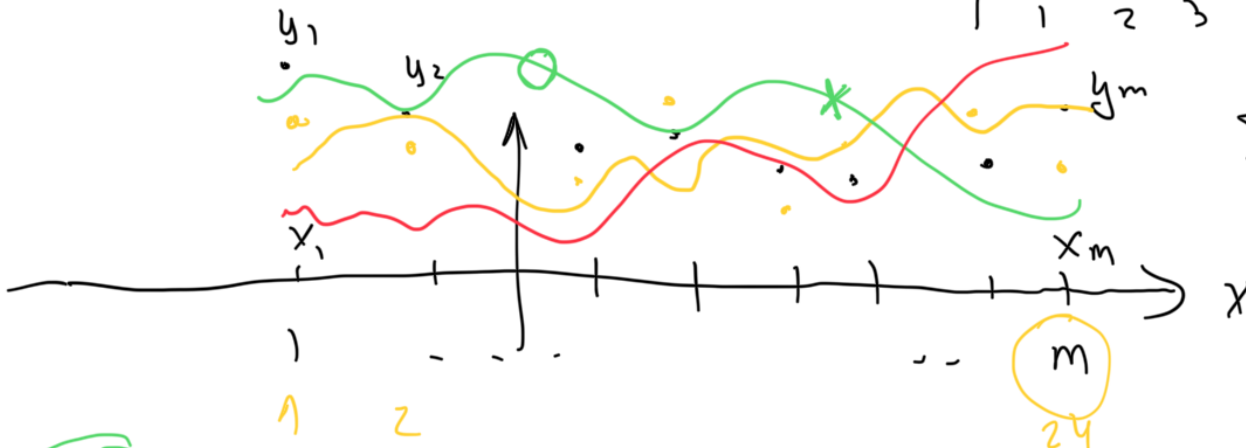
$$N(\vec{0}, \Sigma)$$

$$y(\vec{x}) \sim GP\big(m(\vec{x}), \ k(\vec{x}, \vec{x}')\big)$$

$$\mathrm{cov}\big(y(\vec{x}_1), y(\vec{x}_2)\big) = \mathbb{E}\big(y(\vec{x}_1) - m(\vec{x}_1)\big)\big(y(\vec{x}_2) - m(\vec{x}_2)\big) =$$

$$= k(\vec{x}_1, \vec{x}_2)$$





Σ

$$y(\vec{x}) \sim GP\big(0, \ k(\vec{x}, \vec{x}')\big)$$

$$f(\vec{x}) \in C_1(-\infty, +\infty)$$

$$\forall \varepsilon \ \exists\, y(\vec{x}) : \ \| f(\vec{x}) - y(\vec{x}) \| < \varepsilon$$

$$D = \{(\vec{x}_i, y_i)\}_{i=1}^{m} = (X, \vec{y})$$

$$y_i = f(\vec{x}_i) \qquad y = f(\vec{x})$$

$$\hat{y} = \hat{f}(\vec{x})$$

$$\begin{pmatrix} \vec{y} \\ \hat{y} \end{pmatrix} \sim N\left(\begin{pmatrix} \vec{0} \\ \ \end{pmatrix}, \begin{pmatrix} K & \vec{k} \\ \ & \ \end{pmatrix} \right)$$

$$K = \{ k(\vec{x}_i, \vec{x}_j) \}_{i,j=1}^{m}$$

$$k(\vec{x}_i, \vec{x}_j) = \mathrm{cov}(y_i, y_j)$$

$$\vec{k} = \{ k(\vec{x}, \vec{x}_i) \}_{i=1}^{m}, \quad k(x, \vec{x}_i) = \mathrm{cov}(y, y_i)$$

$$k = k(\vec{x}, \vec{x})$$

$$p(y \mid \vec{y}) = N(\underbrace{\vec{k}^T K^{-1} \hat{\vec{y}}}_{V}, \; k - \vec{k}^T K^{-1} \vec{k})$$

$(\cdot \ldots \cdot) \quad \vec{x} \bullet \quad \bullet$

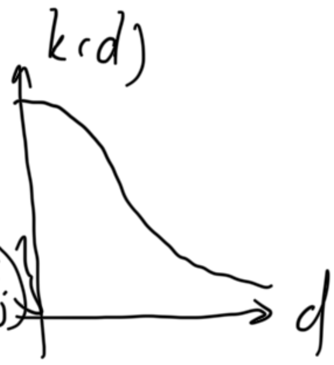
$$\underbrace{\hat{f}(\vec{x}) = \vec{k}^T(\vec{x}) K^{-1} \vec{y}}_{} \quad \text{- prediction}$$

$$= \sum w_i y_i \qquad \vec{w} = \vec{k}^T(\vec{x}) K^{-1}$$

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

$$d^2 = \|\vec{x} - \vec{x}'\|_2^2$$

$$\left[\; k(d) = \exp\left(-\frac{d^2}{\Theta}\right) \right. \qquad K_\Theta = \{ k_\Theta(\vec{x}_i, \vec{x}_j) \}$$



$$\Theta = 1$$

$$p(\vec{y}) = N(\vec{0}, K_\Theta) \qquad \vec{\mu} \qquad (\vec{y} - \vec{\mu})^T \qquad (\vec{y} - \vec{\mu})$$

$$O(m^3) \qquad O(m^3)$$

$$\log p(\vec{y} \mid \Theta) = -\frac{1}{2}\left[\log |K_\Theta| + \vec{y}^T K_\Theta^{-1} \vec{y} \right] \to \max_{\vec{\Theta}} \quad (1)$$

$$p(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2} \frac{(y-\mu)^2}{\sigma^2}\right)$$

$$\log p(y \mid \mu, \sigma^2) = -\frac{1}{2}\left[\log(2\pi) + \log(\sigma^2) + (y-\mu) \frac{1}{\sigma^2} (y-\mu) \right]$$

$$k(\vec{x}, \vec{x}') = \exp\left(-\frac{1}{2} \sum_{i=1}^{d} \frac{(x_i - x_i')^2}{\Theta_i^2}\right)$$

$$\vec{\Theta} = (\Theta_1, \ldots, \Theta_d)$$

Random Fourier Features - an approxim. that works faster

$$N(\underbrace{\vec{k}^T \overbrace{K^{-1} \vec{y}}^{\vec{b}}}_{}, \; k - \vec{k}^T K^{-1} \vec{k}) \qquad k(x_i, x)$$

$$1{\times}m \quad m{\times}m \quad m{\times}1 \qquad \qquad m{\times}m$$

$$\sigma^2 \approx 0 \qquad k(x,x) = \exp\left(-\frac{1}{\Theta}\right) \qquad \exp\left(-\frac{(x_i - x)^2}{\Theta}\right) \sim 0$$

$$O(m^3) \qquad \Downarrow$$

$$O(m): \; \vec{k}^T \vec{b} \qquad O(m^2) \qquad \vec{k} \sim \vec{0}$$