RANDOM FUNCTIONS

$x \in \mathbb{R}^n$, $\mathcal{F} = \{f \mid f(x) = \omega_0 \phi_0(x) + \dots + \omega_{M-1} \phi_{M-1}(x)\}$ is a M-dim vector space

$\phi(x) = (\phi_j(x))$ BASIS FUNCTIONS

What is a distribution over F? $f \Leftrightarrow (\omega_0, \dots, \omega_{M-1}) = \omega$

IT IS equiv. to a distribution over ω vectors!

$\omega \sim \mathcal{N}(0, I)$

$F \ni f$ if we know $\underbrace{p(f(x_1), \dots, f(x_N))}_{\text{FINITE DIMENSIONAL PROJ.}}$, $\forall x_1, \dots, x_N \in \mathbb{R}^n$, then we know the distr. over F.

$\mathcal{F}$

FIX $x \in \mathbb{R}^n$, what is $p(f(x))$. $f(x) \in \mathbb{R}$

$$f(x) = \sum_{j=0}^{M-1} \underset{\mathcal{N}(0,1)}{\omega_j} \underbrace{\phi_j(x)}_{\in \mathbb{R}} = \mathcal{N}(0, \underbrace{\phi^T(x)\, I\, \phi(x)}_{\phi^T(x)\phi(x) = K(x,x)})$$

$F \ni f$ Fix $x_1, x_2 \in \mathbb{R}^n$

$$p(\underbrace{f(x_1), f(x_2)}_{\in \mathbb{R}^2}) = \mathcal{N}\left(0, \begin{pmatrix}\phi(x_1)\\ \phi(x_2)\end{pmatrix}^T I \begin{pmatrix}\phi(x_1)\\ \phi(x_2)\end{pmatrix}\right)$$

$$\begin{pmatrix} f(x_1) \\ f(x_2) \end{pmatrix} = w^T \begin{pmatrix}\phi(x_1)\\ \phi(x_2)\end{pmatrix} = \begin{pmatrix}\phi(x_1) & \phi(x_2)\end{pmatrix} w$$

$$w \sim \mathcal{N}(0, I)$$

$$\operatorname{cov}(f(x_1), f(x_2)) = \begin{pmatrix} \phi^T(x_1)\phi(x_1) & \phi^T(x_1)\phi(x_2) \\ \phi^T(x_2)\phi(x_1) & \phi^T(x_2)\phi(x_2) \end{pmatrix} = \begin{pmatrix} K(x_1,x_1), & K(x_1,x_2) \\ K(x_1,x_2) & K(x_2,x_2) \end{pmatrix}$$

$$p(f(x_1) \dots f(x_n)) = \mathcal{N}(0, K) \qquad K = (K_{ij}) \qquad K_{ij} = K(x_i, x_j)$$

$$\boxed{K(x_1, x_2) = \phi^T(x_1)\phi(x_2)}$$

GAUSSIAN PROCESSES

def: A GP is a stochastic process indexed by a continuous variable $x \in \mathbb{R}^n$ s.t. all finite dimensional distributions are Gaussians.

$f(x)$, $x \in \mathbb{R}^n$ s.t. $\forall x_1, \dots, x_N \in \mathbb{R}^n$, $p(f(x_1), \dots, f(x_N)) = \overset{\text{GAUSSIAN}}{\mathcal{N}(m_N, \Sigma_N)}$

A GP is defined by $\mu : \mathbb{R}^n \to \mathbb{R}$, the mean at x ($\mu(x)$)

$K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, the covariance function between two points.

$f \sim GP(\mu, K)$ iff $\forall x_1, \dots, x_N \in \mathbb{R}^n$

$$p(\underline{f}) = p(f(x_1), \dots, f(x_N)) = \mathcal{N}(\underline{\mu}, \underset{\text{GRAM MATRIX}}{K})$$

$\underline{\mu} = (\mu(x_1), \dots, \mu(x_N))$, $K = (K_{ij})_{ij}$ $K_{ij} = K(x_i, x_j)$

K is such that $\forall x_1, \dots, x_N$, K is positive definite and symmetric.

GP INFERENCE (NOISE FREE CASE)

REGRESSION: $f(x)$ unknown. we observe the true values at points

$\underline{x}: \underbrace{x_1, \dots, x_N}$, so observations are $\underline{f} = \underbrace{f(x_1), \dots, f(x_N)}$

consider TEST POINTS $\underline{x}^* = \underbrace{x_1^*, \dots, x_\nu^*}$, with observations $\underline{f}^* = \underbrace{f(x_1^*) \dots f(x_\nu^*)}$

joint prior distribution of $\underline{f}$ and $\underline{f}^*$. PRIOR IS $GP(0, K)$

$$p\begin{pmatrix} \underline{f} \\ \underline{f}^* \end{pmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(\underline{x}, \underline{x}) & K(\underline{x}, \underline{x}^*) \\ K(\underline{x}^*, \underline{x}) & K(\underline{x}^*, \underline{x}^*) \end{bmatrix}\right)$$

$$K(\underline{x}, \underline{x}^*) = (K_{ij})_{\substack{i=1 \dots N \\ j=1 \dots \nu}} \quad K_{ij} = K(x_i, x_j^*)$$

$$p(\underline{f}^* | \underline{f}) \sim \mathcal{N}\Big(K(\underline{x}^*, \underline{x}) \underbrace{K(\underline{x}, \underline{x})^{-1}} \underline{f}, \; K(\underline{x}^*, \underline{x}^*) - K(\underline{x}^*, \underline{x}) \cdot \underbrace{K(x, x)^{-1}} \cdot K(\underline{x}, x^*) \Big)$$

THIS WORKS FOR ANY $\underline{X}^*$ => THE POSTERIOR IS A GP

GP INFERENCE (NOISE = GAUSSIAN)

$$y(x) = f(x) + \varepsilon \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$\left[\begin{matrix}\text{OBS.} \\ \underline{x}, \underline{y}\end{matrix}\right]$

$$p(\underline{y}) = \mathcal{N}\left(0, K(\underline{x}, \underline{x}) + \sigma^2 I\right)$$

$\underline{x}^*, \underline{f}^* = f(\underline{x}^*)$

$$\Rightarrow p\begin{pmatrix}\underline{y} \\ \underline{f}^*\end{pmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(\underline{x}, \underline{x}) + \sigma^2 I & K(\underline{x}, \underline{x}^*) \\ K(\underline{x}^*, \underline{x}) & K(\underline{x}^*, \underline{x}^*) \end{bmatrix}\right)$$

$$p(\underline{f}^* \mid \underline{y}) \sim \mathcal{N}\Big(K(\underline{x}^*, \underline{x}) \left[K(\underline{x}, \underline{x}) + \sigma^2 I\right]^{-1} \underline{y},$$
$$K(\underline{x}^*, \underline{x}^*) - K(\underline{x}^*, x) \left[K(\underline{x}, \underline{x}) + \sigma^2 I\right]^{-1} K(\underline{x}, \underline{x}^*) \Big)$$

POSTERIOR IS A GP

"cost" is computing $\left(K(\underline{x}, \underline{x}) + \sigma^2 I\right)^{-1}$ this is an $N \times N$ matrix cost $O(N^3)$

PREDICTION x, $f(x)$

$$f(x) \sim \mathcal{N}\left(\boxed{\underline{k}\,(K+\sigma^2 I)^{-1}\,\underline{y}},\ \underline{K(x,x) - \underline{k}^T (K+\sigma^2 I)^{-1} \underline{k}} \right)$$

$$\underline{k} = \left(k(x,\underline{x})\right) = \left(k(x,x_1), \ \ldots\ k(x,x_N)\right)$$

$$\mathbb{E}\left[f(x) \mid \underline{y}\right] = \sum_{i=1}^{N} \alpha_i\, k(x, x_i)$$

$$\alpha = \left[K(\underline{x},\underline{x}) + \sigma^2 I\right]^{-1} \underline{y}$$

GP CLASSIFICATION

2-class problem.

LATENT (NUISANCE) FUNCTION

MODEL $p(C_1|x) = \boxed{\sigma(f(x)) = \pi(x)}$ $f \sim GP(\mu, K)$

OBSERVATIONS : (x_n, y_n) $y_n \in \{0,1\}$ $p(y|f(x)) = \text{Bernoulli}(\sigma(f(x)))$

$(\underline{x}, \underline{y})$ dataset, $f(\underline{x}) = \underline{f}$

N observations $f(x_1), \dots, f(x_N)$

Remark: $\pi(x) = \sigma(f(x))$ is a RANDOM function.

$$p(f(\underline{x})|\underline{y}) = \frac{p(\underline{y}|f(\underline{x}))\, p(f(\underline{x}))}{p(\underline{y})}$$

x^*, $f(x^*)$

$$p(f(x^*)|\underline{y}) = \int p(f(x^*)|f(\underline{x}))\, p(f(\underline{x})|\underline{y})\, df(\underline{x})$$

$$p(\pi(x^*)|\underline{y}) = p(\sigma(f(x^*))|\underline{y})$$

$$\pi^* = \int \sigma(f(x^*))\, p(f(x^*)|\underline{y})\, df(x^*)$$