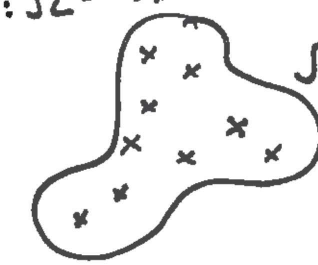
Kernel-based learning

$f: \Omega \to \mathbb{R}$



$\Omega \subseteq \mathbb{R}^d$.

The Mairhuber - Curtis theorem:

Given $\{\phi_1, \dots, \phi_N\}$ basis, $f: \Omega \to \mathbb{R}$, $f(x_1), \dots, f(x_N)$ = samples.

You try $p_N(x) = \sum_{k=1}^{N} c_k \phi_k(x)$ s.t. $p_N(x_k) = f(x_k)$ for $1 \le k \le N$.

Interpolation matrix:

$$\underbrace{\begin{bmatrix} \phi_1(x_1) & \dots & \phi_N(x_1) \\ \vdots & \ddots & \vdots \\ \phi_1(x_N) & \dots & \phi_N(x_N) \end{bmatrix}}_{A} \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_N) \end{bmatrix}$$

If $d \geq 2$ and Ω contains interior pt, A may not be invertible.

Kernel functions:

Bochner, Mercer, etc. (early 1900s)

$K: \Omega \times \Omega \to \mathbb{R}$.

$$p_N(x) = \sum_{k=1}^{N} c_k k(x, x_k), \quad p_N(x_k) = f(x_k), \quad 1 \le k \le N.$$

$$\begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \dots & k(x_N, x_N) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_N) \end{bmatrix}$$

Want to ensure that this matrix is invertible.

Bochner + Mercer: Invertible is hard, but positive def. is much easier.

Design K, so that A is +ve def for any $x_1 \ldots x_N$ in Ω,

Positive-definite kernels

Linear kernels: $\quad k(x,y) = x^T y$

Gaussian $\quad : \quad k(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$.

If H is a Hilbert space, then

$$k(x,y) = (x,y)_H \quad \text{is p.d. kernel.}$$

$$\underline{c}^T A \underline{c} = \sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j (x_i, x_j)_H = \left(\sum_{i=1}^{N} c_i x_i , \sum_{j=1}^{N} c_j x_j \right) \geq 0 .$$

~~Hilbert space → p.d. kernel~~

~~p.d. kernel → Hilbert space~~

Given a kernel, what functions can you learn?

Any function of the form $f(\underline{x}) = \sum_{k=1}^{\infty} c_k K(\underline{x}, \underline{x}_k)$.

$$H_K = \text{Span}\{ k(\cdot, x) \text{ for any } \underline{x} \text{ in } \Omega \}$$

Now, complete to a Hilbert space $\overline{H_K}$.