Kernels

Part 1

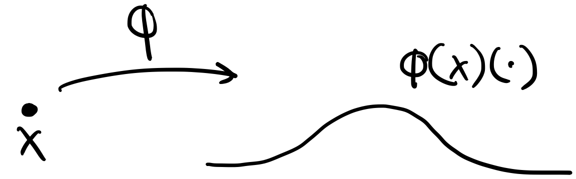
- A $\underline{\text{kernel}}$ is an $\underset{\substack{\uparrow \\ \text{symmetric, bilinear} \\ \text{strictly pos def}}}{\underline{\text{inner product}}}$ for a $\underset{\substack{\uparrow \\ \text{vector space} \\ \text{inner product} \\ \text{complete}}}{\underline{\text{Hilbert space}}}$

- If the world is finite and k is its inner product,
  then k is symmetric and norms ≥ 0 $\rightsquigarrow$ "Gram" matrix $K = \begin{array}{cc} & \ell \\ & \begin{bmatrix} & \\ & k(x_i, x_\ell) \\ & \end{bmatrix} \end{array}$ is symmetric & pos semidef

- Def $\underline{\text{kernel}}$: symmetric & has pos. semidef Gram matrices

Part 2

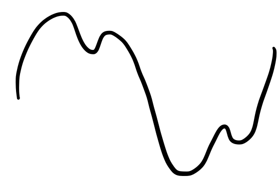
Create an infinite dimensional map ϕ



Use it to create a $\underline{\text{vector space}}$

"vector" $\quad f(\cdot) = \displaystyle\sum_{i=1}^{m} \alpha_i \overbrace{k(\cdot, x_i)}^{\phi}$

"another vector" $\quad g(\cdot) = \displaystyle\sum_{j=1}^{m'} \beta_j \, k(\cdot, x_j')$



with $\underline{\text{inner product}}$ defined $\langle f, g \rangle_{\mathcal{H}_k} := \displaystyle\sum_{i=1}^{m} \sum_{j=1}^{m'} \alpha_i \beta_j \, k(x_i, x_j')$

What if we compute: $\langle k(\cdot, x), f \rangle_{\mathcal{H}_k} = \left\langle k(\cdot, x), \displaystyle\sum_{i=1}^{m} \alpha_i k(\cdot, x_i) \right\rangle_{\mathcal{H}_k}$

$$= \sum_{i=1}^{m} \alpha_i \, k(x, x_i) = f(x)$$

So, $\left.\begin{aligned} \langle k(\cdot, x), f \rangle_{\mathcal{H}_k} &= f(x) \\ \langle k(\cdot, x), k(\cdot, x') \rangle_{\mathcal{H}_k} &= k(x, x') \end{aligned}\right\}$ "reproducing property"

Our vector space is thus a reproducing kernel Hilbert space (RKHS)

$$x \xrightarrow{\phi} k(\cdot, x)$$

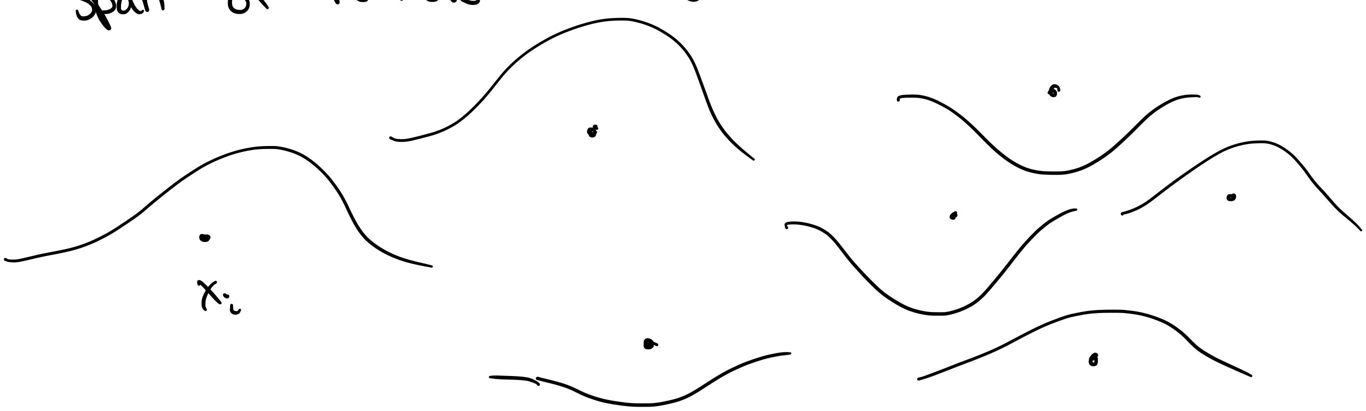
Part 3 Representer Theorem

set X, kernel K, RKHS $\mathcal{H}_K$, loss $\ell : \mathbb{R}^2 \to \mathbb{R}$

Consider $$f^* \in \underset{f \in \mathcal{H}_K}{\operatorname{argmin}} \sum_{i=1}^{n} \ell(f(x_i), y_i) + \|f\|^2_{\mathcal{H}_K}.$$

Theorem $$f^*(\cdot) = \sum_{i=1}^{n} \alpha_i k(x_i, \cdot).$$

Even if RKHS is infinite dimensional, and x_i's are chosen arbitrarily, for any loss ℓ, the solution lies in the span of kernels centered on x_i's !



Part 4 Show that gaussian kernel is a valid kernel.