1.4 Kernels

We'll aim to characterise those similarity measures $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ for which there exists feature map $\phi : \mathcal{X} \to \mathcal{H}$ (where $\mathcal{H}$ is a real inner product space) with

$$k(x, x') = \langle \phi(x), \phi(x') \rangle \quad \forall\, x, x' \in \mathcal{X}$$

Recall that an inner product space $\mathcal{H}$ is a real vector space endowed with a map $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ that obeys

i) symmetry $\langle u, v \rangle = \langle v, u \rangle$

ii) linearity for $a, b \in \mathbb{R}$ $\langle au + bw, v \rangle = a\langle u, v \rangle + b\langle w, v \rangle$

iii) positive definiteness $\langle u, u \rangle \geq 0$ with equality iff $u = 0$

Def^n A <u>positive definite kernel</u> (or simply kernel) k is a symmetric map $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ for which $\forall\, n \in \mathbb{N}$ $x_1, \dots, x_n \in \mathcal{X}$, the matrix $K \in \mathbb{R}^{n \times n}$

$$K_{ij} = k(x_i, x_j) \text{ is positive \underline{semi-definite} } (\alpha^T K \alpha \geq 0).$$

Prop 2 We have a form of the Cauchy-Schwarz inequality:

$$k(x, x')^2 \leq k(x, x)\, k(x', x')$$

Proof: the matrix $\begin{pmatrix} k(x,x) & k(x,x') \\ k(x,x') & k(x',x') \end{pmatrix}$ is pos. semi-def, so its det must be non-negative. $\square$

Prop 3 k defined by ~~def~~ $k(x, x') = \langle \phi(x), \phi(x') \rangle$ is a kernel, where $\phi : \mathcal{X} \to \mathcal{H}$ and $\mathcal{H}$ is an inner product space.

Proof: Let $x_1, \dots, x_n \in \mathcal{X}$, $\alpha \in \mathbb{R}^n$ and consider

$$\sum_{i,j} \alpha_i k(x_i, x_j) \alpha_j = \sum_{i,j} \alpha_i \langle \phi(x_i), \phi(x_j) \rangle \alpha_j = \Big\langle \sum_i \alpha_i \phi(x_i), \sum_j \alpha_j \phi(x_j) \Big\rangle \geq 0$$

$\square$

<u>Examples</u> Prop 4

~~Proof~~ Suppose $k_1, k_2, \dots$ are kernels.

(i) If $\alpha_1, \alpha_2 \geq 0$ then $\alpha_1 k_1 + \alpha_2 k_2$ is a kernel

If $\lim_{m \to \infty} k_m(x, x') = k(x, x')$ exists $\forall\, x, x' \in \mathcal{X}$, then k is a kernel.

(ii) the pointwise product $k = k_1 k_2$ ($k(x, x') = k_1(x, x') k_2(x, x')$) is a kernel

Proof: Example sheet 1 $\square$

Linear kernel $k(x, x') = x^T x'$

Polynomial kernel $k(x, x') = (1 + x^T x')^d$ $d \in \mathbb{N}$ is a kernel.

Note that $1 + x^T x'$ is a kernel owing to the fact that 1 is a kernel and using (i) of prop 4. Then (ii) of prop 4 and induction shows that $(1 + x^T x')^d$ is a kernel $\forall d \in \mathbb{N}$.

Gaussian kernel $k(x, x') = \exp\left(- \frac{\|x - x'\|_2^2}{2\sigma^2}\right)$ σ^2 ← bandwidth

To show this is a kernel $\|x - x'\|_2^2 = \|x\|_2^2 + \|x'\|_2^2 - 2x^T x'$

$$k_1(x, x') = \exp\left(- \frac{\|x\|_2^2}{2\sigma^2}\right) \exp\left(- \frac{\|x'\|_2^2}{2\sigma^2}\right).$$

this is a kernel by prop 3. Next

$$k_2(x, x') = \exp\left(\frac{x^T x'}{\sigma^2}\right) = \sum_{r=0}^{\infty} \frac{(x^T x')^r}{(\sigma^2)^r \, r!}$$ is a kernel by prop 4.

Thus $k = k_1 k_2$ is a kernel by prop 4 (ii).

Sobolev kernel Take $\mathcal{X} = [0, 1]$. $k(x, x') = \min(x, x')$.

this is a kernel since k is the covariance function of Brownian motion.

Jaccard similarity (resemblance) Let $\mathcal{X}$ be the set of all subsets of $\{1, \dots, p\}$.

For $x, x' \in \mathcal{X}$ define $k(x, x') = \begin{cases} \frac{|x \cap x'|}{|x \cup x'|} & \text{if } x \cup x' \neq \emptyset \\ 1 & \text{otherwise} \end{cases}$