

INNER PRODUCT KERNELS & THE KERNEL TRICK

Issues With Nonlinear Feature Maps

Suppose we transform our data via

$$x = \begin{bmatrix} x^{(1)} \\ \vdots \\ x^{(d)} \end{bmatrix} \mapsto \Phi(x) = \begin{bmatrix} \varphi^{(1)}(x) \\ \vdots \\ \varphi^{(m)}(x) \end{bmatrix}, \quad m \gg d$$

where $\varphi^{(j)}$ are nonlinear.

We just discussed how this can lead to ill-conditioned problems, which can be mitigated via regularization.

In addition, since $m \gg d$, there can be an increased _____ burden.

(A)

Example | In ridge regression, we would need to invert an $m \times m$ matrix, or at least solve a linear system with m unknowns.

Fortunately, the following two facts offer a solution:

- Many machine learning algorithms only involve the data through _____.
- For certain feature maps Φ , the function

$$k(x, x') :=$$

has a simple, closed form expression that can be evaluated without explicitly calculating $\Phi(x)$.

Example | Suppose $d=2$ and

$$k(u, v) = (u^T v)^2$$

$$= \left([u^{(1)} \ u^{(2)}] \begin{bmatrix} v^{(1)} \\ v^{(2)} \end{bmatrix} \right)^2$$

$$= (u^{(1)} v^{(1)} + u^{(2)} v^{(2)})^2$$

$$= (u^{(1)})^2 (v^{(1)})^2 + 2u^{(1)} u^{(2)} v^{(1)} v^{(2)} + (u^{(2)})^2 (v^{(2)})^2$$

$$= \langle \Phi(u), \Phi(v) \rangle$$

where

$$\Phi(u) =$$

and the inner-product is the standard dot product.

Now suppose d is arbitrary, and

$$\begin{aligned}k(u, v) &= (u^T v)^2 \\&= \left(\sum_{i=1}^d u^{(i)} v^{(i)} \right)^2 \\&= \left(\sum_{i=1}^d u^{(i)} v^{(i)} \right) \left(\sum_{j=1}^d u^{(j)} v^{(j)} \right) \\&= \sum_{i=1}^d \sum_{j=1}^d u^{(i)} u^{(j)} v^{(i)} v^{(j)}\end{aligned}$$

What is the dimension of the corresponding feature space?

$$\Phi(u) = \left[\underbrace{(u^{(1)})^2, \dots, (u^{(d)})^2}_d, \underbrace{\sqrt{2} u^{(1)} u^{(2)}, \dots, \sqrt{2} u^{(d-1)} u^{(d)}}_{\frac{d(d-1)}{2}} \right]^T$$

Exercise 1 Describe the feature space associated

with

$$k(u, v) = (u^T v)^3$$

when $d = 2$.

Solution |

$$\begin{aligned}k(u, v) &= (u^{(1)} v^{(1)} + u^{(2)} v^{(2)})^3 \\&= (u^{(1)})^3 (v^{(1)})^3 + 3 (u^{(1)})^2 u^{(2)} (v^{(1)})^2 v^{(2)} \\&\quad + 3 u^{(1)} (u^{(2)})^2 v^{(1)} (v^{(2)})^2 + (u^{(2)})^3 (v^{(2)})^3 \\&= \sum_{i=0}^3 \binom{3}{i} (u^{(1)})^{3-i} (u^{(2)})^i \cdot (v^{(1)})^{3-i} (v^{(2)})^i\end{aligned}$$

$$\Rightarrow \Phi(u) = \left[(u^{(1)})^3, \sqrt{3} (u^{(1)})^2 u^{(2)}, \sqrt{3} u^{(1)} (u^{(2)})^2, (u^{(2)})^3 \right]^T$$

More generally,

$$k(u, v) = \left(\sum_{i=1}^d u^{(i)} v^{(i)} \right)^p$$

$$= \sum_{\substack{(j_1, \dots, j_d) \\ \sum j_k = p}} \binom{p}{j_1 \dots j_d} (u^{(1)})^{j_1} \dots (u^{(d)})^{j_d} \cdot (v^{(1)})^{j_1} \dots (v^{(d)})^{j_d}$$

$$\Rightarrow \Phi(u) = \left[\dots, \sqrt{\binom{p}{j_1 \dots j_d}} (u^{(1)})^{j_1} \dots (u^{(d)})^{j_d}, \dots \right]^T$$

\Rightarrow all _____ of degree p .

(D)

PSD kernels

Definition | We say $k(u, v)$ is a _____

_____ kernel if k is (F)

symmetric and, for all n and all

$x_1, \dots, x_n \in \mathbb{R}^d$, the _____ matrix

$$\left[\begin{array}{c} \\ \\ \\ \end{array} \right]_{i,j=1}^n$$

is positive semi-definite.

Theorem | k is an IP kernel \iff

k is a PSD kernel

Proof | Schölkopf and Smola, Learning

with Kernels, or Steinwart and

Christman, Support Vector Machines

Examples of IP kernels

1. Homogeneous polynomial kernel

$$k(u, v) = \langle u; v \rangle^p, \quad p=1, 2, \dots$$

2. Inhomogeneous polynomial kernel

$$k(u, v) = (\langle u; v \rangle + c)^p, \quad p=1, 2, \dots$$

$c > 0$

⑥

$\mathbb{R} \rightarrow$

3. Gaussian kernel

$$k(u, v) = (2\pi\sigma^2)^{-\frac{d}{2}} \exp\left\{-\frac{\|u-v\|^2}{2\sigma^2}\right\}, \quad \sigma > 0$$

→ constant $(2\pi\sigma^2)^{-\frac{d}{2}}$ can be dropped

→ also called radial basis function (RBF) kernel

→ \mathcal{H} is _____ !

⑦

Key

A. computational

B. inner products, $k(x, x') = \langle \Phi(x), \Phi(x') \rangle$

$$C. \left(\sum_{i=1}^2 u^{(i)} v^{(i)} \right)^2 = \left(\sum_{i=1}^2 u^{(i)} v^{(i)} \right) \left(\sum_{j=1}^2 u^{(j)} v^{(j)} \right)$$

$$= \sum_{i=1}^2 \sum_{j=1}^2 u^{(i)} u^{(j)} v^{(i)} v^{(j)}$$

$$= (u^{(1)})^2 (v^{(1)})^2 + 2 u^{(1)} u^{(2)} v^{(1)} v^{(2)} + (u^{(2)})^2 (v^{(2)})^2$$

$$= \langle \Phi(u), \Phi(v) \rangle$$

where

$$\Phi(u) = \begin{bmatrix} (u^{(1)})^2 \\ \sqrt{2} u^{(1)} u^{(2)} \\ (u^{(2)})^2 \end{bmatrix}$$

D. monomials E. inner producting kernel

F. positive semi-definite, gram, $[k(x_i, x_j)]_{i,j=1}^n$

G. all monomials of degree $\leq p$

H. infinite dimensional