# COMP90051 Statistical Machine Learning

Workshop Week 8

**Xudong Han** 

https://github.com/HanXudong/COMP90051 2020 S1

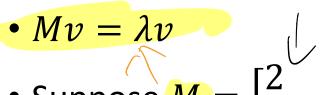
#### Kernel Exercises

- Mercer's Theorem
- Positive Semidefinite/ Positive Definite a symmetric  $n \times n$  matrix M is said to be positive semidefinite if for any n nonzero dim vector z, the scalar  $z^T M z \geq 0$ .

• Eigenvalue  $Mv = \lambda v$  e gen value.

matrix z gen vector

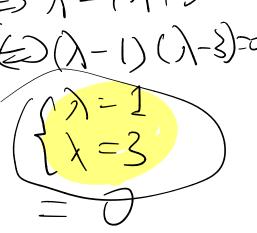




• Suppose 
$$M = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$
  $(2-1)$   $-1 = 0$ 

• How could we compute eigenvalues and eigenvectors?

$$\begin{bmatrix} 2 \\ 1 \\ 2 \\ -\lambda \end{bmatrix}$$



## CKC

Positive Semidefinite/Definite

• For a kernel  $k(x_i, x_j)$ , the full Gram matrix  $(x_i, x_j)$ 

$$= \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix}$$

### Mercer's Theorem

#### Positive Semidefinite

• Given any  $C \in \mathbb{R}^n$ 

• 
$$[c_1 \cdots c_n]$$

$$\begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$
• Since we assume that the kernel  $k(x_i, x_j)$  is valid

kernel

$$C^T K C = \sum_{i,j}^n c_i c_j k(x_i, x_j) \ge 0$$

1-a) 
$$\lambda k(x_i, x_j)$$
 for  $\lambda > 0$ 

• For a kernel  $k'(x_i, x_j)$ , the full Gram matrix

$$K' = \begin{bmatrix} \lambda k(x_1, x_1) & \cdots & \lambda k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \lambda k(x_n, x_1) & \cdots & \lambda k(x_n, x_n) \end{bmatrix}$$

• 
$$C^T K' C = \widehat{\lambda} \sum_{i,j}^n c_i c_j k(x_i, x_j) \ge 0$$

1-b) 
$$k_{\alpha}(x_i, x_j) + k_{\beta}(x_i, x_j) = k(x_i, x_j)$$

• For a kernel  $k'(x_i, x_j)$ , the full Gram matrix

$$\begin{cases} K' \\ = \begin{bmatrix} k_{\alpha}(x_{1}, x_{1}) + k_{\beta}(x_{1}, x_{1}) & \cdots & k_{\alpha}(x_{1}, x_{n}) + k_{\beta}(x_{1}, x_{n}) \\ \vdots & \ddots & \vdots \\ k_{\alpha}(x_{n}, x_{1}) + k_{\beta}(x_{n}, x_{1}) & \cdots & k_{\alpha}(x_{n}, x_{n}) + k_{\beta}(x_{n}, x_{n}) \end{bmatrix}$$

$$\bullet C^{T}K'C = \sum_{i,j}^{n} c_{i}c_{j} \left( k_{\alpha}(x_{i}, x_{j}) + k_{\beta}(x_{i}, x_{j}) \right)$$

$$= \sum_{i,j}^{n} c_{i}c_{j} \left( k_{\alpha}(x_{i}, x_{j}) + k_{\beta}(x_{i}, x_{j}) \right)$$

$$\geq 0$$

1-c) 
$$f(x_i)k(x_i,x_j)f(x_j) = k' \Rightarrow \text{valid}$$

•  $C^T K' C = C^{*T} K C^* \ge 0$ 

CTKC ZO  $C_{\hat{i}}^{*} = C_{\hat{i}} \cdot f(x_{\hat{i}})$ 

Ci St C= Ci f(Xi)

R(Xi, Xj) => valid. C = [0,0,...0]  $C^T \# C > 0$ for any nonzero C

Zij Ci Cj R (Xi, Xj) >0 New bornel R'(Xi, Xj) = f(xi) k(Xi, Xj) f(xj) We want to prove Zij CiGK(Xi) Zo ( ) Zinj Gi G fexi) k(xi, xj) f(xj) > 0  $\int C^{*} = [Gf(x)] Gf(x) \cdots Gnf(xn)]$ Tany value of C. SCXT & CX > 0 }

1-d) 
$$k_{\alpha}(x_i, x_j)k_{\beta}(x_i, x_j) \Rightarrow k'(x_i, x_j)$$

- The Gram matrices  $K_{\alpha}$  and  $K_{\beta}$  are SPSD.
- According to Schur product theorem
   (<a href="https://en.wikipedia.org/wiki/Schur\_product\_theo">https://en.wikipedia.org/wiki/Schur\_product\_theo</a>
   rem), their element-wise product is also PSD

2-b) 
$$\exp(k(x_i, x_j))$$

• 
$$k'(x_i, x_j) = \exp(k(x_i, x_j))$$

#### Taylor series

The Taylor series of a real or complex-valued function f(x) that is infinitely differentiable at a number a is the power series

$$\exp(7) = 1 + x' + \pm x^2 + \cdots$$
 Q 1-d.  
 $(x + y) = 1 + x' + \pm x^2 + \cdots$  Q 1-d.

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x - a)^n \quad \text{for all } x = x \text{ and } x =$$

https://en.wikipedia.org/wiki/Taylor\_series#Exponential\_function

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} (x)^n$$

Thus, 
$$\exp(k(x_i, x_j)) \neq \sum_{n=0}^{\infty} \frac{(k(x_i, x_j))^n}{n!}$$

## Properties of Inner Product

Let u, v, w be vectors and  $\alpha$  be a scalar.

$$\bullet < u + v, w > = < u, w > + < v, w >$$

$$\bullet < \alpha v, w > = \alpha < v, w >$$

$$\bullet < v, w > = < w, v >$$

•  $< v, v > \ge 0$  and equal iff v = 0.

2-c) 
$$\exp\left(-\frac{\left|\left|x-x_{i}\right|\right|^{2}}{2\sigma^{2}}\right)_{val}$$

• 
$$k'(x_i, x_j) = \exp(-2\sigma^{-2} < x - x_i, x - x_i > )$$

$$= \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp\left(\frac{\langle x, x_i \rangle}{\sigma^2}\right) \exp\left(-\frac{x_i^2}{2\sigma^2}\right)$$

• 
$$\exp\left(\frac{\langle x, x_i \rangle}{\sigma^2}\right) = \sum_{n=0}^{\infty} \frac{\left(\frac{\langle x, x_i \rangle}{\sigma^2}\right)^n}{n!} = \sum_{n=0}^{\infty} \frac{\left(\frac{\langle x, x_i \rangle}{\sigma^2}\right)^n}{n!}$$

• Polynomial kernel 
$$k(x, x') = (\langle x, x' \rangle + u)^n$$

$$(\langle \langle \times \times \rangle)^n$$

## 3-a)

$$\langle \mathbf{c}, \mathbf{K} \mathbf{c} \rangle = \sum_{i=1}^{n} c_{i} \sum_{j=1}^{n} \mathbf{K}_{ij} c_{j}$$

$$= \sum_{i} c_{i} \sum_{j} c_{j} \langle \Phi(x_{i}), \Phi(x_{j}) \rangle$$

$$= \sum_{i} c_{i} \left\langle \Phi(x_{i}), \sum_{j} c_{j} \Phi(x_{j}) \right\rangle$$

$$= \left\langle \sum_{i} c_{i} \Phi(x_{i}), \sum_{j} c_{j} \Phi(x_{j}) \right\rangle$$

$$= \left\| \sum_{i} c_{i} \Phi(x_{i}), \right\|^{2} \ge 0 \ \forall \ \mathbf{c} \in \mathbb{R}^{n}$$

## 3-b) Normalised Gaussian Kernel

- Gaussian kernel  $k(x, x') = \exp(\frac{\langle x, x' \rangle}{\sigma^2})$
- Normalised version

$$k'(x, x') = \sqrt{\frac{1}{k(x, x)k(x', x')}}$$

$$= \frac{\exp(\frac{\langle x, x' \rangle}{\sigma^2})}{\sqrt{\exp(\frac{\langle x, x \rangle}{\sigma^2})\exp(\frac{\langle x', x' \rangle}{\sigma^2})}}$$

• See question (2-c)