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 $z^{\intercal} M z \geq 0$.

- Eigenvalue

$$Mv = \lambda v$$

eigen value

matrix

Eigen vector

Eigenvalue

- $Mv = \lambda v$
- Suppose $M = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$
- How could we compute eigenvalues and eigenvectors?

$$\det\left(\cancel{\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}}\right) = 0$$

$$(2-\lambda)^2 - 1 = 0$$

$$\Leftrightarrow \lambda^2 - 4\lambda + 4 - 1 \cancel{= 0}$$

$$\Leftrightarrow \lambda^2 - 4\lambda + 3 = 0$$

$$\Leftrightarrow (\lambda - 1)(\lambda - 3) = 0$$

$$\begin{cases} \lambda = 1 \\ \lambda = 3 \end{cases}$$

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\left[\underbrace{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}}\right] \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \Rightarrow M$$

eigenvalues : 1, 3

$$M v = \lambda v$$

If $\lambda = 1$

if $\lambda = 3$

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 1 \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\begin{bmatrix} 2v_1 + v_2 \\ v_1 + v_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \qquad \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$$

$$\begin{cases} 2v_1 + v_2 = v_1 \\ v_1 + 2v_2 = v_2 \end{cases} \Rightarrow \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$v_1 = -v_2$$

$C^T K C$

Positive Semidefinite/Definite

$C = \begin{bmatrix} a \\ b \end{bmatrix}$ K C PSD

if $a+b=0$

- $\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a+b & a+b \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$

C^T

$$= a^2 + ab + ab + b^2 = (a+b)^2 \geq 0$$

- $\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 + b^2 > 0$ PD

- For a kernel $k(x_i, x_j)$, the full Gram matrix $(X_1, X_2, \cdots X_n)$

valid.

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

Mercer's Theorem – Positive Semidefinite

- Given any $C \in \mathbb{R}^n$
- $\begin{bmatrix} c_1 & \cdots & c_n \end{bmatrix} \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$
- Since we assume that the kernel $k(x_i, x_j)$ is valid kernel

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

valid

$k'(x_i, x_j) = \lambda k(x_i, x_j)$

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

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

k'

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

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

- $k'(x_i, x_j)$ is a valid kernel

$\left[\sum_{i,j} c_i c_j \lambda k(x_i, x_j) \right] \geq 0$

valid

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

$$K' = \begin{bmatrix} k_\alpha(x_1,x_1) + k_\beta(x_1,x_1) & \cdots & k_\alpha(x_1,x_n) + k_\beta(x_1,x_n) \\ \vdots & \ddots & \vdots \\ k_\alpha(x_n,x_1) + k_\beta(x_n,x_1) & \cdots & k_\alpha(x_n,x_n) + k_\beta(x_n,x_n) \end{bmatrix}$$

- $C^T K' C = \sum_{i,j}^{n} c_i c_j \left(k_\alpha(x_i,x_j) + k_\beta(x_i,x_j)\right)$

≥ 0

$$= \underbrace{\sum_{i,j}^{n} c_i c_j k_\alpha(x_i,x_j)}_{\geq 0} + \underbrace{\sum_{i,j}^{n} c_i c_j k_\beta(x_i,x_j)}_{\geq 0}$$

$\Rightarrow K'$ is valid.

$f(x) \Rightarrow$ transformation

1-c) $f(x_i)k(x_i, x_j)f(x_j) = K' \Rightarrow$ valid.

- $C^T K' C = \sum_{i,j}^{n} c_i c_j f(x_i)k(x_i, x_j)f(x_j)$

  $= \sum_{i,j}^{n} [c_i f(x_i)][c_j f(x_j)]\, k(x_i, x_j)$ ← original valid kernel.
- Let $c_i^* = c_i f(x_i)$, $C^{*T} K C^* \geq 0$
- $C^T K' C = C^{*T} K C^* \geq 0$.

$C^T K C \geq 0$

$c_i^* = c_i \cdot f(x_i)$

$c_i \begin{cases} + \\ - \\ 0 \end{cases}$

$c_i^* = c_i\, f(x_i)$

$K(X_i, X_j) \Rightarrow$ valid.

$C \neq [0, 0, \dots 0]$

$C^T K C \geq 0$

for any nonzero C

valid.

$\sum_{i,j}^{n} c_i c_j k(X_i, X_j) \geq 0$

New kernel $k'(X_i, X_j) = f(X_i) k(X_i, X_j) f(X_j)$

We want to prove

$\sum_{i,j}^{n} c_i c_j k'(X_i, X_j) \geq 0$

$\Leftrightarrow \sum_{i,j}^{n} c_i c_j f(X_i) k(X_i, X_j) f(X_j) \geq 0$

$\Leftrightarrow \sum_{ij}^{n} \underbrace{[c_i f(X_i)]}_{C^*} [c_j f(X_j)] k(X_i, X_j) \geq 0$

C^*

$$\begin{cases} C = [c_1 \quad c_2 \quad \cdots \quad c_n]^T \\ C^* = [c_1 f(x_1) \quad c_2 f(x_2) \quad \cdots \quad c_n f(x_n)]^T \end{cases}$$

any value of C.

$\{ C^{*T} K C^* \geq 0 \}$

valid

1-d) $k_\alpha(x_i, x_j) k_\beta(x_i, x_j) \Rightarrow K'(x_i, x_j)$

valid

- The Gram matrices K_α *and* K_β are SPSD.
- According to Schur product theorem (https://en.wikipedia.org/wiki/Schur_product_theorem), their element-wise product is also PSD

$$\begin{bmatrix} k_\alpha(x_1, x_1) & \cdots & k_\alpha(x_1, x_n) \\ \vdots & \ddots & \\ k_\alpha(x_n, x_1) & & k_\alpha(x_n, x_n) \end{bmatrix} \Rightarrow K_\alpha \quad K_\beta$$

$$K' \quad \begin{bmatrix} k_\alpha(x_1, x_1) k_\beta(x_1, x_1) & \cdots & \\ & \ddots & \vdots \\ & \cdots & k_\alpha(x_n, x_n) k_\beta(x_n, x_n) \end{bmatrix}$$

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

$\exp(x) = 1 + x^1 + \frac{1}{2}x^2 + \cdots$

Q 1-d.

$k_\alpha k_\beta \Rightarrow$ valid

$k_\alpha \cdot k_\alpha \Rightarrow$ valid

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

https://en.wikipedia.org/wiki/Taylor_series#Exponential_function

$(k)^n \rightarrow$ valid

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

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

k is valid.

> 0

Properties of Inner Product

Let u, v, w be vectors and α be a scalar.

- $< u + v, w > = < u, w > \ + < v, w >$
- $< \alpha v, w > = \alpha < v, w >$
- $< v, w > = < w, v >$
- $< v, v > \geq 0$ and equal iff $v = 0$.

2-c) $\exp(-\frac{||x-x_i||^2}{2\sigma^2})$ valid.

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

f(x) valid. f(x_i)

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

$\sum_{n=0}^{\infty} \frac{(<x,x_i>)^n}{\sigma^{2n} n!}$

≥ 0

- $\exp\left(-\frac{<x,x_i>}{\sigma^2}\right) = \sum_{n=0}^{\infty}\frac{\left(-\frac{<x,x_i>}{\sigma^2}\right)^n}{n!} = \sum_{n=0}^{\infty}\frac{\left(\frac{<x,x_i>}{\sigma^2}\right)^n}{n!}$

- Polynomial kernel $k(x, x') = (< x, x' > +u)^n$

$(<x, x_i>)^n$.

3-a)

$$
\begin{aligned}
\langle \mathbf{c}, \mathbf{K}\mathbf{c} \rangle &= \sum_{i=1}^{n} c_i \sum_{j=1}^{n} \mathbf{K}_{ij} c_j \\
&= \sum_i c_i \sum_j c_j \left\langle \Phi(x_i), \Phi(x_j) \right\rangle \\
&= \sum_i c_i \left\langle \Phi(x_i), \sum_j c_j \Phi(x_j) \right\rangle \\
&= \left\langle \sum_i c_i \Phi(x_i), \sum_j c_j \Phi(x_j) \right\rangle \\
&= \left\| \sum_i c_i \Phi(x_i), \right\|^2 \geq 0 \; \forall \; \mathbf{c} \in \mathbb{R}^n
\end{aligned}
$$

3-b) Normalised Gaussian Kernel

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

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

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

- See question (2-c)