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 $z^{\top} M z \geq 0$.
- Eigenvalue

  $$Mv = \lambda v$$

Eigenvalue

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

Positive Semidefinite/Definite

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

$$= a^2 + ab + ab + b^2 = (a+b)^2$$

- $\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 + b^2$
- For a kernel $k(x_i, x_j)$, the full Gram matrix

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

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

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

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

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

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

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

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

1-b) $k_\alpha(x_i, x_j) + k_\beta(x_i, x_j)$

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

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

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

1-c) $f(x_i)k(x_i, x_j)f(x_j)$

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

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

- The Gram matrices $K_\alpha\ and\ K_\beta$ are SPSD.
- According to Schur product theorem (https://en.wikipedia.org/wiki/Schur_product_theorem), 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

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

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

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

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

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

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

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

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

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

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)