

CSCI-567: Assignment #3

Due on Friday, October 16, 2015

Saket Choudhary
2170058637

Contents

Problem 4	3
Problem 4: (a)	3
Problem 4: (b)	3
Problem 4: (c)	4
Problem 4: (d)	4
Problem 4: (e)	4
Problem 1	5

Problem 4

Given: $k_1(.,.)$ and $k_2(.,.)$ are kernel function. Thus, for any vector $y \in \mathbf{R}$, $y^T K y \geq 0$ where $K_{ij} = k(x_i, x_j)$ Mercer's theorem requires K to be positive semi-definite.

Problem 4: (a)

$k_3(x, x') = a_1 k_1(x, x') + a_2 k_2(x, x')$ where $a_1, a_2 \geq 0$

Since $k_1(x, x')$ is positive definite, $\forall y \in \mathbf{R}$,

$$y^T K^{(1)} y \geq 0, \quad (4a.1)$$

where

$$K_{ij}^{(1)} = k_1(x_i, x'_j)$$

Similarly,

$$y^T K^{(2)} y \geq 0, \quad (4a.2)$$

where

$$K_{ij}^{(2)} = k_2(x_i, x'_j)$$

Thus, from (4a.1) and (4a.2), we get

$$y^T (K^{(1)} + K^{(2)}) y \geq 0 \quad \forall y \in \mathbf{R} \implies$$

$$y^T K^{(3)} y \geq 0 \quad \forall y \in \mathbf{R}$$

where

$$K_{ij}^{(3)} = k_3(x_i, x'_j)$$

Problem 4: (b)

$k_4(x, x') = f(x)f(x')$ Let $K_{ij}^{(4)} = k_4(x_i, x_j) = f(x_i)f(x'_j)$

Since $f(x)$ is a real valued function, consider $K^{(4)}$

$$K^{(4)} = \begin{bmatrix} f(x_1)f(x'_1) & f(x_1)f(x'_2) & \cdots & f(x_1)f(x'_n) \\ \vdots & & & \\ f(x_n)f(x'_1) & f(x_n)f(x'_2) & \cdots & f(x_n)f(x'_n) \end{bmatrix}$$

$$K^{(4)} = F(\vec{x})_{n \times 1} F(\vec{x})_{1 \times n}^T$$

where

$$F(x)_{1 \times n}^T = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots f(x_n) \end{pmatrix}$$

Now consider $y^T K^{(4)} y = y^T F(x) F(x)^T y = y^T F(x) (y^T F(x))^T = \|y^T F(x)\|_2^2 \geq 0$

Thus, $k_2(.,.)$ is a valid kernel function!.

Problem 4: (c)

$k_5(x, x') = g(k_1(x, x'))$ where g is a polynomial with positive coefficients.

Since g has positive coefficients, $g(x) \geq 0 \forall x \geq 0$

Now consider,

$$y^T K^{(5)} y = (y_1 \ y_2 \ \cdots \ y_n) \times \begin{bmatrix} g(k_1(x_1, x'_1)) & g(k_1(x_1, x'_2)) & \cdots g(k_1(x_1, x'_n)) \\ \vdots & & \\ g(k_1(x_n, x'_1)) & g(k_1(x_n, x'_2)) & \cdots g(k_1(x_n, x'_n)) \end{bmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$y^T K^{(5)} y = y_1 g(k_1(x_1, x'_1)) y_1 + y_2 g(k_1(x_1, x'_2)) y_2 + \cdots y_n g(k_1(x_n, x'_n)) y_n$$

Since $g(k_1(x_i, x_j)) \geq 0$

$$y^T K^{(5)} y \geq 0 \forall y \in \mathbf{R}$$

Thus k_5 is a kernel

Problem 4: (d)

$$k_6(x, x') = k_1(x, x') k_2(x, x')$$

Thus, in terms of our earlier defined matrix notation, $K^{(6)} = K^{(1)} \circ K^{(2)}$ where \circ denotes element wise multiplication (also known as the Hadamard product).

Since, k_1 and k_2 are valid kernel function $\exists v_i w_j$ the eigen vectors of matrix K_1 and K_2 defines such that:

$$K^{(1)} = \sum_i \lambda_i v_i v_i^T \text{ and } K^{(2)} = \sum_j \mu_j w_j w_j^T$$

Now,

$$\begin{aligned} K^{(6)} &= K^{(1)} \circ K^{(2)} \\ &= \sum_i \lambda_i v_i v_i^T \circ \sum_j \mu_j w_j w_j^T \\ &= \sum_{i,j} \lambda_i \mu_j (v_i v_i^T) \circ w_j w_j^T \\ &= \sum_{i,j} \lambda_i \mu_j (v_i \circ w_j)(v_i \circ w_j)^T \\ &\geq 0 \end{aligned}$$

Because $(v_i \circ w_j)(v_i \circ w_j)^T = \|v_i \circ w_j\|_2^2 \geq 0$

Problem 4: (e)

$$k_7(x, x') = \exp(k_1(x, x'))$$

Just like subpart (c), here $g(x) = \exp(x)$ (it's not a polynomial, though that does not affect the derivation we came up with in part (c)). So this is immediate from part (c).

Problem 1

Let $\sigma(a) = \frac{1}{1+e^{-a}}$ and

$$P(Y = 1|X = x) = \sigma(b + w^T x) P(Y = 0|X = x) = 1 - \sigma(b + w^T x)$$

Observe that $Y = 1$ when $b + w^T x \geq 0$ and $Y = 0$ when $b + w^T x < 0$

Thus,

$$\begin{aligned} P(Y = y|X = x) &= \sigma(b + w^T x)^y (1 - \sigma(b + w^T x))^{1-y} \\ \log(P(Y = y|X = x)) &= y \log(\sigma(b + w^T x)) + (1 - y) \log(1 - \sigma(b + w^T x)) \\ &= y \log\left(\frac{\sigma(b + w^T x)}{1 - \sigma(b + w^T x)}\right) + \log(1 - \sigma(b + w^T x)) \\ &= y(b + w^T x) + \log\left(\frac{e^{-(b + w^T x)}}{1 + e^{-(b + w^T x)}}\right) \\ &= y(b + w^T x) + \log\left(\frac{1}{1 + e^{(b + w^T x)}}\right) \\ &= y(b + w^T x) - \log(1 + e^{(b + w^T x)}) \end{aligned} \tag{[1.1]}$$

$$\begin{aligned} \mathcal{L}(w) &= -\log\left(\prod_{i=1}^n P(Y = y_i|X = x_i)\right) \\ &= -\sum_{i=1}^n \log(P(Y = y_i|X = x_i)) \\ &= -\sum_{i=1}^n (y_i(b + w^T x_i) - \log(1 + e^{(b + w^T x_i)})) \\ &= \sum_{i=1}^n (-y_i(b + w^T x_i) + \log(1 + e^{(b + w^T x_i)})) \end{aligned}$$