# CSCI-567: Assignment #3

Due on Friday, October 16, 2015

**Saket Choudhary**
**2170058637**

# Contents

# 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, $\forall y \in \mathbf{R}$,

$$\begin{aligned} y^T K^{(1)} y \geq 0, \\ \text{where} \\ K^{(1)}_{ij} = k_1(x_i, x'_j) \end{aligned} \tag{4a.1}$$

Similarly,

$$\begin{aligned} y^T K^{(2)} y \geq 0, \\ \text{where} \\ K^{(2)}_{ij} = k_2(x_i, x'_j) \end{aligned} \tag{4a.2}$$

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

$$\begin{aligned} y^T (K^{(1)} + K^{(2)}) y \geq 0 \ \forall y \in \mathbf{R} \implies \\ y^T K^{(3)} y \geq 0 \ \forall y \in \mathbf{R} \\ \text{where} \\ K^{(3)}_{ij} = k_3(x_i, x'_j) \end{aligned}$$

## Problem 4: (b)

$k_4(x, x') = f(x)f(x')$ Let $K^{(4)}_{ij} = k_4(x_i, x_j) = f(x_i)f(x'_j)$
Since $f(x)$ is a real valued function, consider $K^{(4)}$

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

$$\begin{aligned} K^{(4)} = F(\vec{x})_{n\times 1} F(\vec{x})^T_{1\times n} \\ \text{where} \\ F(x)^T_{1\times n} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots f(x_n) \end{pmatrix} \end{aligned}$$

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 continued on next page...

## 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: $K^{(1)} = \sum_i \lambda_i v_i v_i^T$ and $K^{(2)} = \sum_j \mu_j w_j {w_j}^T$
Now,

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

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

## 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

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

Observe that $Y=1$ when $b+w^Tx \geq 0$ and $Y=0$ when $b+w^Tx < 0$
Thus,

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

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