Prof. Dr. Rudolf Mathar, Dr. Arash Behboodi, Markus Rothe

Exercise 12
- Proposed Solution -

Friday, January 25, 2019

Solution of Problem 1

a) b is given as $-\frac{1}{2}\mathbf{a}^T(\mathbf{x}_1 + \mathbf{x}_2) = 3$.

b) Supporting vectors are those with $\lambda_i \neq 0$, which are $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5$.

$\mathbf{x}_i$	y_i	λ_i	$\mathbf{x}_i$	y_i	λ_i
$\mathbf{x}_1 = \begin{pmatrix}1\\1\end{pmatrix}$	$y_1 = -1$	$\lambda_1 = 0$	$\mathbf{x}_4 = \begin{pmatrix}0.5\\-0.5\end{pmatrix}$	$y_4 = 1$	$\lambda_4 = 4.73$
$\mathbf{x}_2 = \begin{pmatrix}2\\0\end{pmatrix}$	$y_2 = -1$	$\lambda_2 = 0.67$	$\mathbf{x}_5 = \begin{pmatrix}-2\\1\end{pmatrix}$	$y_5 = 1$	$\lambda_5 = 0.94$
$\mathbf{x}_3 = \begin{pmatrix}0\\0\end{pmatrix}$	$y_3 = -1$	$\lambda_3 = 5$	$\mathbf{x}_6 = \begin{pmatrix}0\\-1\end{pmatrix}$	$y_6 = 1$	$\lambda_1 = 0$

c) For those vectors, the normal vector of the hyperplane is obtained as:

$$\mathbf{a} = \sum_{i=1}^{6} \lambda_i y_i \mathbf{x}_i = \lambda_2 y_2 \mathbf{x}_2 + \lambda_3 y_3 \mathbf{x}_3 + \lambda_4 y_4 \mathbf{x}_4 + \lambda_5 y_5 \mathbf{x}_5$$

$$\mathbf{a} = -0.67\begin{pmatrix}2\\0\end{pmatrix} - 5\begin{pmatrix}0\\0\end{pmatrix} + 4.73\begin{pmatrix}0.5\\-0.5\end{pmatrix} + 0.94\begin{pmatrix}-2\\1\end{pmatrix} = \begin{pmatrix}-0.86\\-1.43\end{pmatrix}$$

To find b, take two support vectors $\mathbf{x}_k$ and $\mathbf{x}_l$ with $y_k = 1$ and $y_l = -1$ with $0 < \lambda < 5$. For these support vectors, we have $y_i(\mathbf{a}^T\mathbf{x}_i + b) = 1$. Hence:

$$b^\star = \frac{-1}{2}\mathbf{a}^{\star T}(\mathbf{x}_k + \mathbf{x}_l) = -\frac{1}{2}\begin{pmatrix}-0.86 & -1.43\end{pmatrix}\left(\begin{pmatrix}2\\0\end{pmatrix} + \begin{pmatrix}-2\\1\end{pmatrix}\right) = \frac{1.43}{2} = 0.715. \quad (1)$$

Solution of Problem 2

From Mercer's Theorem we know that a kernel $K : \mathbb{R}^p \to \mathbb{R}$ is a valid kernel if and only if for any $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ the kernel matrix $\mathbf{K} = (K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1,\ldots,n}$ is non-negative definite.

a) For $K(\mathbf{x}, \mathbf{y}) = K_1(\mathbf{x}, \mathbf{y})$ we have that

$$\mathbf{K} = (K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1,\ldots,n} = (\alpha K_1(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1,\ldots,n} = \alpha \mathbf{K}_1 .$$

Then, since $\mathbf{K}_1$ is non-negative definite and $\alpha > 0$, it holds

$$\mathbf{z}^T\mathbf{K}\mathbf{z} = \underbrace{\alpha}_{>0}\underbrace{z^T\mathbf{K}\mathbf{z}}_{\geq 0} \geq 0\,, \quad \forall \mathbf{z} \in \mathbb{R}^p \quad \Rightarrow \quad \mathbf{K}\text{is non-negative definite.}$$

b) For $K(\mathbf{x},\mathbf{y}) = K_1(\mathbf{x},\mathbf{y}) + K_2(\mathbf{x},\mathbf{y})$ we have that $\mathbf{K} = \mathbf{K}_1 + \mathbf{K}_2$, thus

$$\mathbf{z}^T\mathbf{K}\mathbf{z} == \mathbf{z}^T(\mathbf{K}_1+\mathbf{K}_2)\mathbf{z} = \underbrace{z^T\mathbf{K}_1\mathbf{z}}_{\geq 0} + \underbrace{z^T\mathbf{K}_2\mathbf{z}}_{\geq 0} \geq 0\,, \quad \forall \mathbf{z} \in \mathbb{R}^p \Rightarrow \mathbf{K}\text{is non-negative definite.}$$

c) For $K(x,y) = K_1(x,y)K_2(x,y)$ we have that $\mathbf{K} = K_1 \odot \mathbf{K}2$, where $\odot$ denotes the-Hadamard product (i.e., point-wise multiplication) between two matrices. Let $\mathbf{K}_1 = \sum_{i=1}^n \lambda_i \mathbf{v}_i\mathbf{v}_i^T$ and $\mathbf{K}_2 = \sum_{i=1}^n \rho_i \mathbf{u}_i\mathbf{u}_i^T$ be the spectral decompositions of $\mathbf{K}_1$ and $\mathbf{K}_2$ respectively. Note that, since these matrices non-negative definite we have that their eigenvalues λ_i and ρ_i are non-negative for all $i = 1,\dots,n$. Therefore, we get

$$K = (\sum_{i=1}^n \lambda_i \mathbf{v}_i\mathbf{v}_i^T)\odot(\sum_{j=1}^n \rho_j \mathbf{u}_j\mathbf{u}_j^T) = \sum_{i=1}^n\sum_{j=1}^n \rho_i\lambda_j \mathbf{v}_i\mathbf{v}_i^T \odot \mathbf{u}_j\mathbf{u}_j^T = \sum_{i=1}^n\sum_{j=1}^n \rho_i\lambda_j(\mathbf{v}_i\odot\mathbf{u}_j)(\mathbf{v}_i\odot\mathbf{u}_j)^T\,.$$

Note that for any $i,j = 1,\dots,n$ the matrix $\rho_j\lambda_i(\mathbf{v}_i \odot \mathbf{u}_j)(\mathbf{v}_i \odot \mathbf{u}_j)^T$ is a rank-1 matrix with eigenvalue $\rho_j\lambda_i \geq 0$, thus a non-negative definite matrix. This means that $\mathbf{K}$ is a sum of non-negative definite matrices, therefore, as shown in **b)**, it is also non-negative definite.

Solution of Problem 3

If a Kernel is given by $K(\mathbf{x},\mathbf{z}) = (\mathbf{x}^T\mathbf{z} + c)^d$, we have:

$$\begin{aligned}(\mathbf{x}^T\mathbf{z} + c)^d &= (\sum_{i=1}^n x_i z_i + c)^d = \sum \beta(\alpha_1,\dots,\alpha_{n+1})(x_1z_1)^{\alpha_1}\dots(x_nz_n)^{\alpha_n}c^{\alpha_{n+1}} \\ &= \sum \left(\sqrt{c^{\alpha_{n+1}}\beta(\alpha_1,\dots,\alpha_{n+1})}x_1^{\alpha_1}\dots x_n^{\alpha_n}\right)\left(\sqrt{c^{\alpha_{n+1}}\beta(\alpha_1,\dots,\alpha_{n+1})}z_1^{\alpha_1}\dots z_n^{\alpha_n}\right)\end{aligned}$$

where the sum is taken over all $(\alpha_1,\dots,\alpha_{n+1}) \in \mathbb{N}^{n+1}$ such that $\sum_{i=1}^{n+1}\alpha_i = d, \alpha_i \in \mathbb{N}$. The number of all these monomials are given by the number of answers to the above equation which is $\binom{n+d}{d}$. Therefore the feature map can be considered as:

$$\phi(\mathbf{x}) = \left(\sqrt{c^{\alpha_{n+1}}\beta(\alpha_1,\dots,\alpha_{n+1})}x_1^{\alpha_1}\dots x_n^{\alpha_n}\right)_{\sum_{i=1}^{n+1}\alpha_i = d, \alpha_i \in \mathbb{N}} \in \mathbb{R}^{\binom{n+d}{d}}.$$