Modern Multivariate Statistical Techniques

Jinhong Du,15338039

November 26, 2018

Content

1 Ex 11.7

Let $z \in \mathbb{R}$ and define the $(2m+1)$-dimensional Φ-mapping,

$$\Phi(z) = (2^{-\frac{1}{2}}, \cos z, \cdots, \cos mz, \sin z, \cdots, \sin mz)^\top$$

Using this mapping, show that the kernel $K(x,y) = \langle \Phi(x), \Phi(y) \rangle, x, y \in \mathbb{R}$, reduces to the Dirichlet kernel given by

$$K(x,y) = \frac{\sin\left[\left(m+\frac{1}{2}\right)\delta\right]}{2\sin\frac{\delta}{2}}$$

where $\delta = x - y$.

Proof.

$$\begin{aligned}
K(x,y) &= \langle \Phi(x), \Phi(y) \rangle \\
&= \langle (2^{-\frac{1}{2}}, \cos x, \cdots, \cos mx, \sin x, \cdots, \sin mx)^\top, (2^{-\frac{1}{2}}, \cos y, \cdots, \cos my, \sin y, \cdots, \sin my)^\top \rangle \\
&= \frac{1}{2} + \sum_{j=1}^{m} [\cos(jx)\cos(jy) + \sin(jx)\sin(jy)] \\
&= \frac{1}{2} + \sum_{j=1}^{m} \cos[j(x-y)] \\
&= \frac{1}{\sin\frac{\delta}{2}} \sum_{j=0}^{m} \sin\frac{\delta}{2}\cos(j\delta) - \frac{1}{2} \\
&= \frac{1}{2\sin\frac{\delta}{2}} \sum_{j=0}^{m} \left\{ \sin\left[\left(j+1-\frac{1}{2}\right)\delta\right] - \sin\left[\left(j-\frac{1}{2}\right)\delta\right] \right\} - \frac{1}{2} \\
&= \frac{\sin\left[\left(m+\frac{1}{2}\right)\delta\right] + \sin\left(\frac{1}{2}\delta\right)]}{2\sin\frac{\delta}{2}} - \frac{1}{2} \\
&= \frac{\sin\left[\left(m+\frac{1}{2}\right)\delta\right]}{2\sin\frac{\delta}{2}}
\end{aligned}$$

□

2 Ex 11.8

Show that the homogeneous polynomial kernel, $K(x,y) = \langle x, y \rangle^d$, satisfies Mercer's condition (11.54).

$$\mathbf{K} = \begin{bmatrix} K(\mathbf{x}_1,\mathbf{x}_1) & K(\mathbf{x}_1,\mathbf{x}_2) & \cdots & K(\mathbf{x}_1,\mathbf{x}_n) \\ K(\mathbf{x}_2,\mathbf{x}_1) & K(\mathbf{x}_2,\mathbf{x}_2) & \cdots & K(\mathbf{x}_2,\mathbf{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ K(\mathbf{x}_n,\mathbf{x}_1) & K(\mathbf{x}_n,\mathbf{x}_2) & \cdots & K(\mathbf{x}_n,\mathbf{x}_n) \end{bmatrix}$$

is non-negative-definite.

Proof. We first prove the product of two kernel is still a kernel, i.e., let $\mathbf{K}_1, \mathbf{K}_2$ be the Gram matrix for two kernels, then $\mathbf{K}'$ given by $K(x,y) = K_1(x,y)K_2(x,y)$ is positive semi-definite.

The eigen-decomposition of $\mathbf{K}_j$ is given by

$$\mathbf{K}_j = \sum_{i=1}^{n} \lambda_i^{(j)} \mathbf{u}_i^{(j)\top} \mathbf{u}_i^{(j)}$$

where $\lambda_i^{(j)} \geq 0$ $(i = 1, 2, \cdots, n)$.

Therefore,

$$\begin{aligned}
\mathbf{K}' &= \mathbf{K}_1 \odot \mathbf{K}_2 \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i^{(1)} \lambda_j^{(2)} (\mathbf{u}_i^{(1)\top} \mathbf{u}_i^{(1)}) \odot (\mathbf{u}_j^{(2)\top} \mathbf{u}_j^{(2)}) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i^{(1)} \lambda_j^{(2)} (\mathbf{u}_i^{(1)} \odot \mathbf{u}_j^{(2)}) (\mathbf{u}_i^{(1)} \odot \mathbf{u}_j^{(2)})^\top
\end{aligned}$$

$\forall\ \mathbf{a} \in \mathbb{R}^n$,

$$\begin{aligned}
\mathbf{a}^\top \mathbf{K}' \mathbf{a} &= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i^{(1)} \lambda_j^{(2)} \mathbf{a}^\top (\mathbf{u}_i^{(1)} \odot \mathbf{u}_j^{(2)}) (\mathbf{u}_i^{(1)} \odot \mathbf{u}_j^{(2)})^\top \mathbf{a} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i^{(1)} \lambda_j^{(2)} \| (\mathbf{u}_i^{(1)} \odot \mathbf{u}_j^{(2)})^\top \mathbf{a} \|_2^2 \\
&\geq 0
\end{aligned}$$

Therefore, $\mathbf{K}'$ is positive semi-definite.

Since $K(x,y) = \langle x, y \rangle^d = \langle x, y \rangle \times \cdots \times \langle x, y \rangle$, it follows that $\mathbf{K}$ is positive semi-definite. □

3 Ex 11.2

In the support vector regression problem using a quadratic ε-insensitive loss function, formulate and solve the resulting optimization problem.

A quadratic ε-insensitive loss function is given by

$$L_2(y, \mu(\mathbf{x}), \varepsilon) = \max\{0, [y - \mu(\mathbf{x})]^2 - \varepsilon\}$$

where $\varepsilon > 0$.

For quadratic ε-insensitive loss, the primal optimization problem is to find $\beta_0, \boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\boldsymbol{\xi}'$ to

$$\begin{aligned}
\min \quad & \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \frac{C}{2} \sum_{i=1}^{n} (\xi_i^2 + \xi_i'^2) \\
s.t. \quad & y_i - (\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}) \leq \varepsilon + \xi_i'
\end{aligned}$$

$$(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}) - y_i \leq \varepsilon + \xi_i$$

where $C > 0$ is a constant.

The primal function is given by

$$\begin{aligned}
F_P(\beta_0, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\xi}', \mathbf{a}, \mathbf{b}) = \frac{1}{2}\|\beta\|^2 + \frac{C}{2}\sum_{i=1}^{n}(\xi_i^2 + \xi_i'^2) \\
+ \sum_{i=1}^{n} a_i[y_i - (\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}) - \varepsilon - \xi_i'] \\
+ \sum_{i=1}^{n} b_i[(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}) - y_i - \varepsilon - \xi_i]
\end{aligned}$$

Setting the derivatives to 0,

$$\begin{aligned}
\frac{\partial F_P}{\partial \beta_0} &= \sum_{i=1}^{n}(a_i - b_i) = 0 \\
\frac{\partial F_P}{\partial \boldsymbol{\beta}} &= \boldsymbol{\beta} - \sum_{i=1}^{n} a_i \mathbf{x}_i + \sum_{i=1}^{n} b_i \mathbf{x}_i = 0 \\
\frac{\partial F_P}{\partial \boldsymbol{\xi}'} &= C\boldsymbol{\xi}' - \mathbf{a} = 0 \\
\frac{\partial F_P}{\partial \boldsymbol{\xi}} &= C\boldsymbol{\xi} - \mathbf{b} = 0
\end{aligned}$$

A stationary solution yields,

$$\begin{aligned}
\boldsymbol{\beta}^* &= \sum_{i=1}^{n}(a_i - b_i)\mathbf{x}_i \\
\sum_{i=1}^{n}(a_i - b_i) &= 0 \\
\boldsymbol{\xi}' &= \frac{1}{C}\mathbf{a} \\
\boldsymbol{\xi} &= \frac{1}{C}\mathbf{b}
\end{aligned}$$

Substituting the solution into the primal function gives us the dual function

$$\begin{aligned}
F_D(\mathbf{a}, \mathbf{b}) &= \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \frac{1}{2C}(\mathbf{a}^\top\mathbf{a} + \mathbf{b}^\top\mathbf{b}) \\
&\quad + (\mathbf{a} - \mathbf{b})\mathbf{y} - \sum_{i=1}^{n}(a_i - b_i)\mathbf{x}_i^\top \boldsymbol{\beta} - \sum_{i=1}^{n}(a_i + b_i)\varepsilon - \frac{1}{C}[\mathbf{a}^\top\mathbf{a} + \mathbf{b}^\top\mathbf{b}] \\
&= \frac{1}{2}\|\boldsymbol{\beta}\|^2 - \|\boldsymbol{\beta}\|^2 \\
&\quad - \frac{1}{C}[\mathbf{a}^\top\mathbf{a} + \mathbf{b}^\top\mathbf{b}] + (\mathbf{a} - \mathbf{b})\mathbf{y} - \varepsilon(\mathbf{a} + \mathbf{b})^\top \mathbf{1} \\
&= -\frac{1}{2}\|\boldsymbol{\beta}\|^2 - \frac{1}{C}[\mathbf{a}^\top\mathbf{a} + \mathbf{b}^\top\mathbf{b}] + (\mathbf{a} - \mathbf{b})\mathbf{y} - \varepsilon(\mathbf{a} + \mathbf{b})^\top \mathbf{1} \\
&= -\frac{1}{2}(\mathbf{a} - \mathbf{b})^\top \mathbf{K}(\mathbf{a} - \mathbf{b}) - \frac{1}{C}[\mathbf{a}^\top\mathbf{a} + \mathbf{b}^\top\mathbf{b}] + (\mathbf{a} - \mathbf{b})\mathbf{y} - \varepsilon(\mathbf{a} + \mathbf{b})^\top \mathbf{1}
\end{aligned}$$

where

$$\mathbf{K} = \left(\langle \mathbf{x}_i, \mathbf{x}_j \rangle\right)_{1 \le i,j \le n}$$

Therefore, the dual problem is given by

$$\begin{aligned} \max \quad & P_D \\ s.t. \quad & \mathbf{a}, \mathbf{b} \succeq \mathbf{0} \\ & (\mathbf{a} - \mathbf{b})^\top \mathbf{1} = 0 \end{aligned}$$

From the KKT conditions, for $i = 1, 2, \cdots, n$,

$$\begin{aligned} a_i \left[y_i - (\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}) - \varepsilon - \frac{a_i}{C} \right] &= 0 \\ b_i \left[(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}) - y_i - \varepsilon - \frac{b_i}{C} \right] &= 0 \\ a_i b_i &= 0 \end{aligned}$$

solve them for $\mathbf{a}$ and $\mathbf{b}$. If $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ are the solution, then

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \sum_{i=1}^{n} (\hat{a}_i - \hat{b}_i) \mathbf{x}_i \\ \hat{\beta}_0 &= \frac{1}{|\{i | a_i > 0\}| + |\{i | b_i > 0\}|} \left[\sum_{\{i | \hat{a}_i > 0\}} \left(y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} - \varepsilon - \frac{\hat{a}_i}{C} \right) + \sum_{\{i | \hat{b}_i > 0\}} \left(y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + \varepsilon + \frac{\hat{b}_i}{C} \right) \right] \end{aligned}$$