ML HW3

SNIR HORDAN 205689581, ALEXANDER SHENDER 328626114

1. Let $\hat{w}$ =V$(\Sigma^{+})^{2}\Sigma^{T}U^{T}y$ where $\Sigma\in\mathbb{R}^{m\times d}$ with the singular values derived from the SVD, as the entries on the diagonal , and U,V square orthogonal matrices of order $m\times m$ and $d\times d$, respectly.

Notice that

$$X\hat{w}=U\Sigma\underbrace{V^{T}V}_{=I_{d}}(\Sigma^{+})^{2}\Sigma^{T}U^{T}y=$$
$$=U\underbrace{\Sigma(\Sigma^{+})^{2}\Sigma^{T}}_{see\,(1)}U^{T}y=UU^{T}y=y$$

(1) $\Sigma=\begin{pmatrix}\sigma_{1} & 0 & 0 & 0 & 0 & 0\\ 0 & \sigma_{2} & 0 & 0 & 0 & 0\\ 0 & 0 & . & 0 & 0 & 0\\ 0 & 0 & 0 & \sigma_{r} & 0 & 0\end{pmatrix}$ and $\Sigma^{+}=\begin{pmatrix}\frac{1}{\sigma_{1}} & 0 & 0 & 0\\ 0 & . & 0 & 0\\ 0 & 0 & . & 0\\ 0 & 0 & 0 & \frac{1}{\sigma_{r}}\end{pmatrix}$ then $\Sigma\Sigma^{+}=\begin{pmatrix}1 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & . & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0\end{pmatrix}$ Then $\Sigma(\Sigma^{+})^{2}\Sigma^{T}=I_{m}$

Then $||\,X\hat{w}-y\,||_{2}^{2}=0$

The metric $||\cdot||_{2}:\mathbb{R}^{d}\rightarrow\mathbb{R}^{+}$is non-negative then $\hat{w}\in argmin_{w}||\,Xw-y\,||_{2}^{2}$

2. .1.1.

Define $\phi_{3}:\mathbb{R}^{d}\rightarrow\mathbb{R}^{2d}$, $\phi_{3}\left(\vec{x}\right)=\begin{pmatrix}| \\ \phi_{1}\left(\vec{x}\right)\\ | \\ | \\ \phi_{2}\left(\vec{x}\right)\\ |\end{pmatrix}$

Define $K_{3}:\mathbb{R}\times\mathbb{R}\rightarrow\mathbb{R}$ $K_{3}\left(x,x'\right)=\left\langle \phi_{3}\left(x\right),\phi_{3}\left(x'\right)\right\rangle$

Then

$$K_{3}\left(x,x'\right)=\left\langle \phi_{3}\left(x\right),\phi_{3}\left(x'\right)\right\rangle =\left\langle \begin{pmatrix}| \\ \phi_{1}\left(x\right)\\ | \\ | \\ \phi_{2}\left(x\right)\\ |\end{pmatrix},\begin{pmatrix}| \\ \phi_{1}\left(x'\right)\\ | \\ | \\ \phi_{2}\left(x'\right)\\ |\end{pmatrix}\right\rangle =\left\langle \begin{pmatrix}| \\ \phi_{1}\left(x\right)\\ |\end{pmatrix},\begin{pmatrix}| \\ \phi_{1}\left(x'\right)\\ |\end{pmatrix}\right\rangle +\left\langle \begin{pmatrix}| \\ \phi_{2}\left(x\right)\\ |\end{pmatrix},\begin{pmatrix}| \\ \phi_{2}\left(x'\right)\\ |\end{pmatrix}\right\rangle =$$
$$=K_{1}\left(x,x'\right)+K_{2}\left(x,x'\right)$$

K_{3} is a valid kernel function iff its Gram matrix is Positive Semi-Definite.

Let $0_{m}\neq z\in\mathbb{R}^{m}$ and $x_{1},..,x_{m}\in\mathbb{R}^{d}$

Then $z^{T}G_{K_{3}}z=z^{T}(G_{K_{1}}+G_{K_{2}})z=z^{T}G_{K_{!}}z+z^{T}G_{K_{2}}z\geq0$, because K_{1},K_{2} are valid kernel functions.

Therefore K_{3} is a valid kernel function.

2.1.2

Define $\phi_{4}:\mathbb{R}\rightarrow\mathbb{R}$, $\phi_{4}\left(\vec{x}\right)=f\left(\vec{x}\right)\phi_{1}\left(\vec{x}\right)$ and let $f\,:\mathbb{R}^{d}\rightarrow\mathbb{R}$ be a function.

Define $K_{4}:\mathbb{R}\times\mathbb{R}\rightarrow\mathbb{R}$ $K_{4}\left(x,x'\right)=\left\langle \phi_{4}\left(x\right),\phi_{4}\left(x'\right)\right\rangle$

Then $K_{4}\left(x,x'\right)=\left\langle \phi_{4}\left(x\right),\phi_{4}\left(x'\right)\right\rangle =\left\langle f\left(x\right)\phi_{1}\left(x\right),\,f\left(x'\right)\phi_{1}\left(x'\right)\right\rangle =f\left(x\right)f\left(x'\right)\left\langle \phi_{1}\left(x\right),\phi_{1}\left(x'\right)\right\rangle =f\left(x\right)f\left(x'\right)K_{1}\left(x,x'\right)$

K_{4} is a valid kernel function iff its Gram matrix is Positive Semi-Definite.

Let $0_{m}\neq z\in\mathbb{R}^{m}$ and $x_{1},..,x_{m}\in\mathbb{R}^{d}$

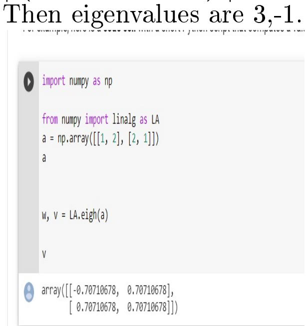
Then $z^{T}G_{K_{4}}z=z^{T}diag\left(f\left(x_{1}\right),...,f\left(x_{m}\right)\right)^{T}G_{K_{1}}diag\left(f\left(x_{1}\right),...,f\left(x_{m}\right)\right)z\geq0$ because $diag\left(f\left(x_{1}\right),...,f\left(x_{m}\right)\right)z\in\mathbb{R}$ and $G_{K_{1}}$ is PSD because K_{1} is a valid kernel function.

Therefore K_4 is a valid kernel function.

2.2.1
Observe $\left(\begin{array}{cc} K(x_1,x_1) & K(x_1,x_2) \\ K(x_2,x_1) & K(x_2,x_2) \end{array}\right) = \left(\begin{array}{cc} 1 & 2 \\ 2 & 1 \end{array}\right)$

$\left|\left(\begin{array}{cc} 1-\lambda & 2 \\ 2 & 1-\lambda \end{array}\right)\right| = (1-\lambda)^2 - 4 = -3 - 2\lambda + \lambda^2 = (\lambda-3)(\lambda+1)$

Then eigenvalues are 3,-1.



```
import numpy as np

from numpy import linalg as LA
a = np.array([[1, 2], [2, 1]])
a



w, v = LA.eigh(a)


v
```

```
array([[-0.70710678,  0.70710678],
       [ 0.70710678,  0.70710678]])
```

Above are the eigenvectors of norm 1.

2.2.2
Define $\hat{v} = \left(\begin{array}{c} -1 \\ 1 \end{array}\right)$. v is an eigevector of the Gram matrix. Note, $y_1 = 1, y_2 = -1$

Let $\lambda \geq 0$.

$L_{SVM}(c\hat{v}) = \lambda \left(\begin{array}{cc} -c & c \end{array}\right) \left(\begin{array}{cc} 1 & 2 \\ 2 & 1 \end{array}\right) \left(\begin{array}{c} -c \\ c \end{array}\right) + \frac{1}{2}\sum_{i=1}^{2} \max\left\{0, 1 - y_i (cGv)_i\right\}$

$= -\lambda 2c^2 + \frac{1}{2}\left[\max\{0, 1-1(+c)\} + \max\{0, 1-(-1)(-c)\}\right] = ...$

$... = -\lambda 2c^2 + \frac{1}{2}[2+2c] = -\lambda c^2 + 1 - c \underset{c\to\infty}{\to} -\infty$

Non-PSD kernels are "problematic" because there exists an eigenvector with a negative eigenvalue. Then, exists a contradiction to the existance of a minimal value over the set $\{L_{SVM}(\alpha) \mid \alpha \in \mathbb{R}^m\}$.

2.2.3. Assume $G \geq 0$ and let $\lambda \geq 0$.
2.2.3.1. Note $G \geq 0$ if $\forall z \neq 0_m$, $z^T G z \geq 0$

$\frac{1}{m}\sum\limits_{i=1}^{m} \max\left\{0, 1 - y_i (G\alpha)_i\right\} \geq 0$ by definition of the hinge-loss function.

$\lambda \alpha^T G \alpha \geq 0$ because G is PSD.

Then $L_{SVM}(\alpha) = \min_\alpha \lambda \alpha^T G \alpha + \frac{1}{m}\sum\limits_{i=1}^{m} \max\left\{0, 1 - y_i (G\alpha)_i\right\}$

2.2.3.2.
By theorem optimal solution $\hat{w} = \sum\limits_{i=1}^{m} \alpha_i x_i$.

Let $w = \sum\limits_{i=1}^{m} \alpha_i x_i$ be a separating hyperplane.

By the Representor Theorem, if w is an optimal solution to the Soft-SVM optimization problem then $y_i \langle w, x_i \rangle = \sum\limits_{j=1}^{m} \alpha_j \langle x_i, x_j \rangle \geq 1$ Then $\alpha_i = 0$.

Then we can assume above claim holds for w, because we are attempting to minimize the loss function over $\alpha \in \mathbb{R}^m$.

Therefore,

$L_{SVM}(\alpha) = \lambda \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{m} \alpha_i \alpha_j \langle x_i, x_j \rangle + \frac{1}{m}\sum\limits_{i=1}^{m} \max\left\{0, 1 - \sum\limits_{j=1}^{m} \alpha_j \langle x_i, x_j \rangle\right\}$

$$\underset{Representor\,Thm.}{=} \lambda \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i\alpha_j \langle x_i, x_j\rangle + 1 - \frac{1}{m} \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i\alpha_j \langle x_i, x_j\rangle = 1 + \left(\lambda - \frac{1}{m}\right)\alpha^T G\alpha$$

Note $G \geq 0$ if $\forall z \neq 0_m$, $z^T G z \geq 0$

Then $L_{SVM}(\alpha) = 1 + \left(\lambda - \frac{1}{m}\right)\alpha^T G\alpha \implies \min_\alpha L_{SVM}(\alpha) \leq 1$ because $||\alpha||_2$ can be arbitrarily small.