CSCI 567 (ML)
BHASKAR GOYAL
USC ID - 6547367383

PROBLEM 1.

1.1 We are given,

$$f(w) = \sum_{i=1}^{N} y_i w^T x_i \,,$$

and are asked to maximize the $f(w)$ given $\|w\|_2 \leq 1$.

Objective function,

$$\max_{w} \left(\sum_{i=1}^{N} y_i w^T x_i \right)$$

subject to

$$w^T w \leq 1$$

For the constraint minimization problem, writing the above objective function in standard form:

$$\min_{w} \left(-\sum_{i=1}^{N} y_i w^T x_i \right)$$

subject to,

$$w^T w \cancel{\leq} -1 \leq 0$$

To solve, we will use the Lagrangian.

Let the lagrangian multiplier be λ,

$$L(w, \lambda) = -\sum_{i=1}^{N} y_i w^T x_i + \lambda (w^T w - 1)$$

Now, we need to find.

$$\min_{w} L(w, \lambda).$$

For this we will use the stationary condition.

$$\frac{\partial}{\partial w} L(w, \lambda) = 0.$$

$$\frac{\partial}{\partial w} L(w, \lambda) = -\sum_{i=1}^{N} y_i x_i + (2w)\lambda = 0$$

$$\Rightarrow -\sum_{i=1}^{N} y_i x_i + 2\lambda w = 0$$

$$\Rightarrow \quad 2\lambda w = +\sum_{i=1}^{N} y_i x_i$$

$$\boxed{w^* = +\frac{1}{2\lambda} \sum_{i=1}^{N} y_i x_i}$$

1.2 We apply a transformation, $\phi : \mathcal{X} \to K$.

$$K(x, x') = \phi(x)^T \phi(x')$$

Constraint minimization problem,

$$\min \left(- \sum_{i=1}^{N} y_i w^T \phi(x_i) \right)$$

subject to,

$$w^T w - 1 \leq 0$$

optimal w^* for this problem,

$$\boxed{w^* = \frac{1}{2\lambda} \sum_{i=1}^{N} y_i \, \phi(x_i)}$$

1.3 To find the Dual form of $f(w)$, we first have to calculate $L(w, \lambda)$, then minimize it for w, find optimal value of w^*, and then finally, maximize λ over $\min_{w} L(w^*, \lambda)$.

$$L(w, \lambda) = - \sum_{i=1}^{N} y_i w^T \phi(x_i) + \lambda (w^T w - 1).$$

But we need to minimize $L(w, \lambda)$ over w,

$$\min_{w} L(w, \lambda)$$

And we know, $w^* = \frac{1}{2\lambda} \sum_{i=1}^{N} y_i \phi(x_i)$.

Subsituting the value of w, in $L(w, \lambda)$.

$$\Rightarrow \min_{w} L(w^*, \lambda) =$$

$$-\sum_{i=1}^{N} y_i \left(\frac{1}{2\lambda} \sum_{j=1}^{N} y_j \phi(x_j) \right)^T \phi(x_i)$$

$$+ \lambda \left[\left(\frac{1}{2\lambda} \sum_{i=1}^{N} y_i \phi(x_i) \right)^T \left(\frac{1}{2\lambda} \sum_{j=1}^{N} y_j \phi(x_j) \right) - 1 \right]$$

To find the Dual form, maximize over λ,

$$\Rightarrow \max_{\lambda} \min_{w} L(w, \lambda) =$$

$$\max_{\lambda} \left(-\sum_{i=1}^{N} y_i \left(\frac{1}{2\lambda} \sum_{j=1}^{N} y_j \phi(x_j) \right)^T \phi(x_i) + \lambda \left[\left(\frac{1}{2\lambda} \sum_{i=1}^{N} y_i \phi(x_i) \right)^T \left(\frac{1}{2\lambda} \sum_{j=1}^{N} y_j \phi(x_j) \right) - 1 \right] \right)$$

Subject to,

$$\lambda \geq 0.$$

Simplifying

$$\Rightarrow L = -\frac{1}{2\lambda} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \phi^T(x_i)\, \phi(x_j)$$

$$+ \frac{1}{4\lambda} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i \phi^T(x_i)\, y_j \phi(x_j) - \lambda$$

$$L = -\frac{1}{4\lambda} \sum_{i,j} y_i y_j \phi^T(x_i)\, \phi(x_j) - \lambda.$$

Therefore, the dual formulation is,

$$\Rightarrow \max_{\lambda} \min_{w} L(w, \lambda) =$$

$$\boxed{\max_{\lambda} \left[-\frac{1}{4\lambda} \sum_{i,j} y_i y_j \phi^T(x_i)\, \phi(x_j) - \lambda \right] \quad \text{subject to, } \lambda \geq 0}$$

1.4 Yes, the optimization's dual problem can be kernalized because we have a term of $\phi^T(x)\phi(x')$ in the objective function.

$$L = -\frac{1}{4\lambda} \sum_{i,j} y_i y_j \phi^T(x_i)\, \phi(x_j) - \lambda$$

$$= -\frac{1}{4\lambda} \sum_{i,j} y_i y_j K(x_i, x_j) - \lambda$$

$$\left\{ \text{where } \phi^T(x_i)\phi(x_j) = K(x_i, x_j) \right\}$$

Therefore, the new objective function is,

$$\max_{\lambda} \min_{w} L(w, \lambda)$$

$$= \max_{\{\lambda\}} \left[\frac{-1}{4\lambda} \sum_{i,j} y_i y_j K(x_i, x_j) - \lambda \right]$$

subject to $\lambda \geq 0$.

Prediction Rule

Yes, the prediction rule can also be kernalized as

$$\text{prediction} = \text{sign}\left(w^T \phi(x) \right)$$

Put w^* in prediction rule,

$$= \text{sign}\left(\left(\frac{1}{2\lambda} \sum_{i=1}^{N} y_i \phi(x_i) \right)^T \phi(x) \right)$$

$$= \text{sign}\left(\frac{1}{2\lambda} \sum_{i=1}^{N} y_i \phi^T(x_i) \phi(x) \right)$$

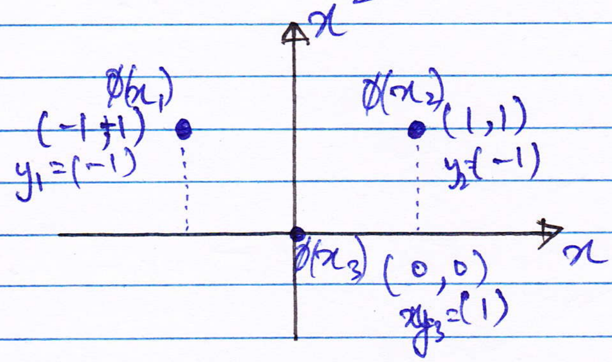
$$= \text{sign}\left(\frac{1}{2\lambda} \sum_{i=1}^{N} y_i K(x_i, x) \right)$$

$\therefore$, we can kernalize the prediction rule as we were able to substitute $\phi^T(x_i) \phi(x)$ with $K(x_i, x)$ in the prediction rule.

PROBLEM 2

(2.1) **No**, the three points in one-dimensional feature space, can not be perfectly separated with a linear separator.
This is because if we pick any point in the 1-D feature space, one side of the data points will be classified as +1 and other will be classified as −1. Therefore, there will always be one data point which is misclassified.
∴ we cannot perfectly separate the data points with a linear separator.

(2.2) $\phi(x) = [x, x^2]^T$



$$\phi(-1) = [-1, 1]^T$$
$$\phi(0) = [0, 0]^T$$
$$\phi(1) = [1, 1]^T$$

Yes, there exists a linear decision boundary. that can separate this two dimensional feature space.
One example of such decision boundary is

$x^2 = 0.25$. or $x^2 = 0.5$
(2nd dimension of $\phi(x)$

2.3

$$K(x,x') = \phi^T(x)\,\phi(x')$$
$$= \begin{bmatrix} x & x^2 \end{bmatrix} \begin{bmatrix} x' \\ x'^2 \end{bmatrix}$$
$$= xx' + (xx')^2$$
$$= xx'(1 + xx')$$

K (gram matrix)

$$K = \begin{bmatrix} \phi^T(x_1)\,\phi(x_1) & \phi^T(x_2)\,\phi(x_1) & \phi^T(x_3)\,\phi(x_1) \\ \phi^T(x_1)\,\phi(x_2) & \phi^T(x_2)\,\phi(x_2) & \phi^T(x_3)\,\phi(x_2) \\ \phi^T(x_1)\,\phi(x_3) & \phi^T(x_2)\,\phi(x_3) & \phi^T(x_3)\,\phi(x_3) \end{bmatrix}$$

$$= \begin{bmatrix} K(x_1,x_1) & K(x_2,x_1) & K(x_3,x_1) \\ K(x_1,x_2) & K(x_2,x_2) & K(x_3,x_2) \\ K(x_1,x_3) & K(x_2,x_3) & K(x_3,x_3) \end{bmatrix}$$

$$= \begin{bmatrix} -1.-1(1+(-1)^2) & 1.-1(1+-1) & 0.-1(1+0) \\ -1.1(1-1) & 1.1(1+1) & 0.1(1+0) \\ -1.0(1+0) & 1.0(1+0) & 0.0(1+0) \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

To verify that K is a positive semidefinite, we need to show that

$$u^T K u \geq 0. \quad \forall u \in R^{3\times 1}$$

let u be $[a \; b \; c]^T$, so,

$\Rightarrow \quad u^T K u$

$$= [a \; b \; c] \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

$$= [a \; b \; c] \begin{bmatrix} 2a \\ 2b \\ 0 \end{bmatrix}$$

$$= 2a^2 + 2b^2 \geq 0.$$

Because a^2 and b^2 are positive ~~const~~,

$\therefore \quad u^T K u \geq 0$,

Hence, K is a positive semidefinite matrix.

2.4 Dual of the SVM is given by,

$$\max_{\{\alpha_n\}} \sum_n \alpha_n - \frac{1}{2} \sum_{m,n} y_m y_n \alpha_m \alpha_n \phi^T(x_m) \phi(x_n$$

$$\text{s.t.} \quad \sum_n \alpha_n y_n = 0.$$

$$0 \leq \alpha_n \leq C \quad \forall n.$$

let α be $[\alpha_1 \; \alpha_2 \; \alpha_3]^T$

then, plugging in the data points, we get,

$$\Rightarrow \max_{\{\alpha_1,\alpha_2,\alpha_3\}} \sum_{i=1}^{3} \alpha_i - \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} y_i y_j \alpha_i \alpha_j \underbrace{\phi^T(x_i)\phi(x_j)}_{=K(x_i,\, x_j)}$$

s.t

$$\alpha_1 y_1 + \alpha_2 y_2 + y_3\alpha_3 = 0$$

and

$$0 \le \alpha_n \le C \quad \forall n \in \{1,2,3\}$$

Using, K gram matrix values, we get:

$$\Rightarrow \max_{\{\alpha_1,\alpha_2,\alpha_3\}} \alpha_1+\alpha_2+\alpha_3 - \frac{1}{2}\Big(\alpha_1\alpha_1 y_1 y_1 \phi^T(x_1)\phi(x_1) + \alpha_2\alpha_2 y_2 y_2 \phi^T(x_2)\phi(x_2)\Big)$$

$$\text{s.t} \quad (-1)\alpha_1 + (-1)\alpha_2 + (1)(\alpha_3) = 0.$$
$$0 \le \alpha_n \le C \quad \forall n \in \{1,2,3\}.$$

Plugging, $y_1 = -1$, $y_2 = -1$, $y_3 = 1$.

$$\Rightarrow \max_{\{\alpha_1,\alpha_2,\alpha_3\}} \alpha_1+\alpha_2+\alpha_3 - \frac{1}{2}\left(2\alpha_1^2 + 2\alpha_2^2\right)$$

$$\text{St.} \quad \alpha_1+\alpha_2 = \alpha_3$$
$$0 \le \alpha_n \le C \quad \forall n \in \{1,2,3\}.$$

$\Rightarrow$ Using constraint $\alpha_1+\alpha_2 = \alpha_3$, we get.

$$\max_{\{\alpha_1,\alpha_2,\alpha_3\}} 2\alpha_1 + 2\alpha_2 - (\alpha_1^2+\alpha_2^2)$$

$$\text{s.t} \quad 0 \le \alpha_n \le C \quad \forall n \in \{1,2,3\}$$

$$\boxed{\begin{array}{l} \Rightarrow \max\limits_{\{\alpha_1,\alpha_2,\alpha_3\}} 2\alpha_1 - \alpha_1^2 + 2\alpha_2 - \alpha_2^2 \\ \text{s.t.} \quad 0 \le \alpha_1 \le C \\ \qquad 0 \le \alpha_2 \le C \\ \qquad 0 \le \alpha_3 \le C \end{array}}$$

(C can be further simplified if $C = +\infty$)

For simplicity we are not using C, therefore, the values became.

$$\boxed{\begin{array}{ll} \max\limits_{\{\alpha_1, \alpha_2, \alpha_3\}} & 2\alpha_1 - \alpha_1^2 + 2\alpha_2 - \alpha_2^2 \\ \text{s.t} & \\ & \alpha_1 \geq 0 \\ & \alpha_2 \geq 0 \\ & \alpha_3 \geq 0 \\ & \alpha_1 + \alpha_2 = \alpha_3 \end{array} \quad \left\{ \begin{array}{l} \text{for} \\ \text{simplicity} \\ \text{leaving} \\ \text{C out} \end{array} \right\}}$$

Hence, above is the final dual form of the SVM, given.

2.5

Solving the dual form,
Clearly, we need to maximize α_1, α_2 separately,
Hence, we will get,

$$\alpha_1^* = 1 \qquad \left(\frac{\partial f}{\partial \alpha_1} = 2 - 2\alpha_1 = 0 \Rightarrow \alpha_1 = 1\right)$$
$$\alpha_2^* = 1 \qquad \left(\frac{\partial f}{\partial \alpha_2} = 2 - 2\alpha_2 = 0 \Rightarrow \alpha_2 = 1\right)$$

and we know $\alpha_1 + \alpha_2 = \alpha_3$, so,
$\alpha_3^* = 2$.

$$\boxed{\alpha^* = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}}$$

Now, that we have value of α^*, we can get w^* and b^*.

$$w^* = \sum_n \alpha_n^* y_n \phi(x_n)$$
$$= \alpha_1^* y_1 \phi(x_1) + \alpha_2^* y_2 \phi(x_2) + \alpha_3^* y_3 \phi(x_3)$$
$$= (1)(-1)\begin{bmatrix} -1 \\ 1 \end{bmatrix} + (1)(-1)\begin{bmatrix} 1 \\ 1 \end{bmatrix} + (2)(1)\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

$$\boxed{w^* = \begin{bmatrix} 0 \\ -2 \end{bmatrix}}$$

$$b^* = y_n - w^{*T} \phi(x_n)$$
Plugging $x_1 = -1$ and $y_1 = -1$,

$$b^* = -1 - \begin{bmatrix} 0 \\ -2 \end{bmatrix}^T \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$= -1 - \begin{bmatrix} 0 & -2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$= -1 - (-2)$$

$$\boxed{b^* = 1}$$

$\therefore$, after solving dual form analytically, we get.

$$w^* = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

$$b^* = 1$$

PROBLEM 3.

3.1

$$X = \begin{bmatrix} 4 & 1 \\ 2 & 3 \\ 5 & 4 \\ 1 & 0 \end{bmatrix}$$

First, we need to center the data points.

The mean of first column of X.

$$= \frac{4+2+5+1}{4} = \frac{12}{4} = 3.$$

The mean of second column of X

$$= \frac{1+3+4+0}{4} = \frac{8}{4} = 2$$

$$\underset{(\text{centered})}{X'} = \begin{bmatrix} 4-3 & 1-2 \\ 2-3 & 3-2 \\ 5-3 & 4-2 \\ 1-3 & 0-2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 2 & 2 \\ -2 & -2 \end{bmatrix}$$

$$\underset{(\text{centered})}{X'} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 2 & 2 \\ -2 & -2 \end{bmatrix}$$

Now, we will find covariance matrix $X'^T X'$

$$X'^T X' = \begin{bmatrix} 1 & -1 & 2 & -2 \\ -1 & 1 & 2 & -2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 2 & 2 \\ -2 & -2 \end{bmatrix}$$

$$x'^T x' = \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix}$$

$\Rightarrow$ Computing the eigen values and eigen vector,

$$\det(x'^T x' - \lambda I) = 0$$

$$\det\left(\begin{bmatrix} 10-\lambda & 6 \\ 6 & 10-\lambda \end{bmatrix}\right) = 0$$

$$(10-\lambda)^2 - 6^2 = 0$$
$$(10-\lambda-6)(10-\lambda+6) = 0$$
$$(\lambda-4)(\lambda-16) = 0.$$

$\underline{\lambda = 16, 4.}$

$\Rightarrow$ for $\lambda = 16$, eigen vector is,

$$\Rightarrow x'^T x' v = \lambda v$$
$$\Rightarrow \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = 16\begin{bmatrix} a \\ b \end{bmatrix}$$

$\Rightarrow$ $a = b$, let $a = b = 1$.

so, unit length. eigen vector is,

$$\underline{V_1 = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]^T \text{ for } \lambda_1 = 16}$$

$\Rightarrow$ for $\lambda = 4$, eigen vector is,

$$\Rightarrow x'^T x' v = \lambda v$$
$$\Rightarrow \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = 4\begin{bmatrix} a \\ b \end{bmatrix}$$

$\Rightarrow$ $a = -b$ let $a = 1, \Rightarrow b = -1$

so, unit length eigen vector is.

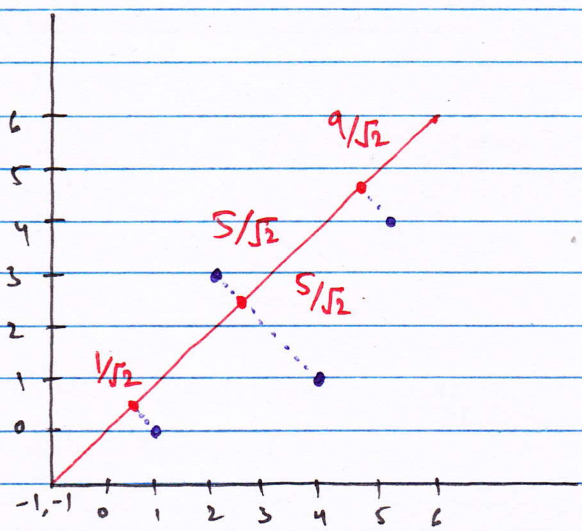
$$\underline{V_2 = \left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right]^T \text{ for } \lambda_2 = 4}$$

We will choose the larger eigen value's eigen vector

$\therefore$, the unit-length principal component direction of X,

$$\boxed{v_1 = \left(1/\sqrt{2} \quad 1/\sqrt{2} \right)^T \qquad \text{for } \lambda = 16}$$

3.2



For point $(1,0)$, project point $= v \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}}(1) + \frac{1}{\sqrt{2}}(0)$

$$= \frac{1}{\sqrt{2}}$$

For point $(2,3)$, project point $= v \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \frac{1}{\sqrt{2}}(2) + \frac{1}{\sqrt{2}}(3)$

$$= \frac{5}{\sqrt{2}}.$$

for point (4,1),

projected point $= v.\begin{bmatrix}4\\1\end{bmatrix}$

$$= \frac{1}{\sqrt{2}}(4) + \frac{1}{\sqrt{2}}(1)$$

$$= \frac{5}{\sqrt{2}}$$

for point (5,4)

projected point $= v.\begin{bmatrix}5\\4\end{bmatrix}$

$$= \frac{1}{\sqrt{2}}(5) + \frac{1}{\sqrt{2}}(4)$$

$$= \frac{9}{\sqrt{2}}.$$

$\therefore$, the projected point coordinates are.

$(1,0) \Rightarrow 1/\sqrt{2}$

$(4,1) \Rightarrow 5/\sqrt{2}$

$(2,3) \Rightarrow 5/\sqrt{2}$

$(5,4) \Rightarrow 9/\sqrt{2}$.