# CMPE 257 Machine Learning Spring 2019
## HW#3

Name: Akshata Deo
Sjsu id: 012565761

---

**1.** Consider a dataset with 3 data points in 2-D plane:
X = [[0,0],[0,-1],[-2,0]]
Y = [[-1],[-1],[+1]]

Now, the constraint optimization we are given is as follows-
Min(b,w): (½)$w^Tw$
Subject to: $y_n(w^Tx_n + b) >= 1$
w belongs to $R^2$
b belongs to R
In 2D, a hyper plane is specified by the parameters (b,w1,w2). The inequality on a particular row is the separability constraint for the corresponding data point in that row.

-b >= 1 ----- eq 1
$w_2 – b >= 1$ ----- eq 2
$-2w_1 + b >= 1$ ----- eq 3
Combining eq (1 & 3), we get-
$-2 w_1 >= 2$ => $w_1 <= -1$
1st equation gives b <= -1
from eq 2, we get-
$w_2 <= 0$

This means that (½) $(w_1^2 + w_2^2)$ is minimum when w1 = -1 and w2 = 0. One can easily verify that ($b^*$ = −1,$w_1$* = -1,$w_2$* = 0) satisfies all 3 constraints, minimizes (½) $(w_1^2 + w_2^2)$,
Therefore the optimal hyperplane is-
g(x) = sign ($-x_1 + x_2 – 1$)
Margin = 1/||w*|| => 1

---

**2.** Dual problem from above-
X = [[0,0],[2,2],[2,0]]
Y = [-1, -1, +1]
W = [1.2 , -3.2]
b = -0.5
setting up the dual-
min of α $(1/2)\alpha^TG\alpha – 1^T\alpha$
subject to: $y^T\alpha = 0$ , α.>=0
where, G = $XsXs^T$

Xs = [[0,0],[-2,-2],[2,0]]
So, $XsXs^T$ = [[0,0,0],[0,8,4],[0,-4,4]]
min of α (1/2) * [α1, α2, α3] * [[0,0,0],[0,8,4],[0,-4,4]] * $[α1, α2, α3]^T$ – [1,1,1] * [α1, α2, α3]
min of α (1/2) *[0, $8α_2$-$4α_3$, $-4α_2$+$4α_3$] * $[α1, α2, α3]^T$ – (α1+α2+α3)

min of $\alpha$ $[0, 4\alpha_2-2\alpha_3, -2\alpha_2+2\alpha_3] * [\alpha1, \alpha2, \alpha3]^T - \alpha1-\alpha2-\alpha3$

min of $\alpha$ $(4\alpha_2)^2 -2\alpha_2\alpha_3 -2\alpha_2\alpha_3 + (2\alpha_3)^2 - \alpha1-\alpha2-\alpha3$

min of $\alpha$ $(4\alpha_2)^2 + (2\alpha_3)^2 -4\alpha_2\alpha_3 - \alpha1-\alpha2-\alpha3$ ----- 1st equation

constraints: $-\alpha1-\alpha2+\alpha3 = 0$ $\alpha1>=0, \alpha2>=0, \alpha3>=0$

substitute into 1st equation, $\alpha1=\alpha3-\alpha2$

objective: min of $\alpha$ $(4\alpha_2)^2 + (2\alpha_3)^2 -4\alpha_2\alpha_3 -2\alpha3$

delta J = 0

so, $[[dJ/d\alpha1],[ dJ/d\alpha2],[ dJ/d\alpha3]] = 0$

$[0, 8\alpha_2-4\alpha_3, 4\alpha_3-4\alpha_2-2] = 0$

$8\alpha_2-4\alpha_3 = 0$

$2\alpha_2 = \alpha_3$

Substitute, $4\alpha_3-4\alpha_2-2 = 0$

$4(2\alpha_2)-4\alpha_2-2 = 0$

$\alpha_2 = ½$ , $\alpha_3 = 1$, $\alpha_1 = ½$

$w^* = \sum_{n=1to3} y_n\alpha_n x_n$

$= -1(1/2)[[0],[0]] + (-1)(1/2)[[2],[2]] + 1(1)[[2],[0]]$

$= [[-1+2],[-1]]$

$= [[-1],[+1]]$

$b^* = -1$

---

**3.**

**(a). & (b).** File attached

**(c). & (d).** File attached

**(d).** As shown in the figure, if we take random hypotheses for large data points (such as 1 million as provided in question), the alpha will be evenly distributed in the region [-1,1]

For SVM the ‘a’ will not be evenly distributed and create a curve with maxima at point 0. The number of ‘a’ will decrease if we move in either direction.

---

**4.**

**(a).** File attached

**(b).** $\Phi$3 appears to be over fitted.

**(c).** File attached

**(d).** File attached

---

**Refrences:**


- Book forum: http://book.caltech.edu/bookforum/
- Learning from data e-book.