AML Assignment - 2

Subhani shaik
MT18117

(1)

(1) Given data points -ve class $\Rightarrow$ $a=(-1,0)$, $b=(2,-2)$ $\Rightarrow$ (-1) = -ve class
+ve class $\Rightarrow$ $c=(1,0)$ $\Rightarrow$ $+1$. class

Given seperators

(i) (a) $x_1 + x_2 = 0$ as Decision boundary.

generally For decision boundary $w^T x + b = 0$

Here $w_1 x_1 + w_2 x_2 + b = 0$. $\Rightarrow [w_1, w_2] = [1 \ 1]$

$\rightarrow$ check for data points that these weights are classifying correctly or not & check for max margin property $\rho = \frac{2}{\|w\|}$

$w^T x + b \geq 1 \Rightarrow$ +ve
$w^T x + b \leq -1 \Rightarrow$ -ve

-ve For $a = (-1, 0)$ $\Rightarrow$ $1(-1) + 1(0) = -1 < 0$. $\rightarrow$ classified correctly

-ve for $b = (2, -2)$ $\Rightarrow$ $1(2) + 1(-2) = 0$ $\rightarrow$ class on Decision boundary.

+ve for $c = (1, 0)$ $\Rightarrow$ $1(1) + 1(0) = 1$ $\rightarrow$ classified correctly.

$\Rightarrow$ point 'b' is on Decision boundary (Seperator)

$\rightarrow$ It is not maximizing margin width. (misclassification).

$\Rightarrow$ Not satisfying svm conditions.

(b) $x_1 + 1.5 x_2 = 0$. $\Rightarrow [w_1, w_2] = (1, 1.5)$

$\Rightarrow$ $a = (-1, 0)$ $\Rightarrow$ $1(-1) + 1.5(0) = -1$ $\Rightarrow$ ✓

$b = (2, -2)$ $\Rightarrow$ $1(2) + 1.5(-2) = 2 - 3 = -1$ $\Rightarrow$ ✓

$C = (1, 0)$ $\Rightarrow$ $1(1) + 1.5(0) = 1$ $\Rightarrow$ ✓

all are classified correctly. & all a, b, c are Support vectors Here

check for max margin $\rho = \frac{2}{\|w\|}$.

$\|w\| = \sqrt{1^2 + 1.5^2} = \sqrt{1 + 2.25} = \sqrt{3.35} = 1.830 \approx 2$

$\sqrt{4} = \pm 2$

$\Rightarrow \rho = \frac{2}{\sqrt{3.35}} = \frac{2}{1.830} \approx 1.$

Here all are support vectors $\Rightarrow$ distance from decision boundary to all points should be same (as Linear svm Decision boundary will be parallel to all support vectors).

→ calculating distance to 'a' $=(-1,0)$, $(w_1,w_2)=(1,1.5)$

$= \frac{|-1|}{\|w\|} = 0.5$

$b = (1,0) = \frac{1}{\|w\|} = 0.5$

$c = (2,-2) = 0.5.$

all distances are equal ⇒ It satisfied all svm conditions.

(c) $x_1 + 2x_2 = 0$

$(w_1, w_2) = (1,2)$

for -ve $a = (-1,0)$ ⇒ $-1+0 = -1$ ⇒ ✓

+ve $b = (1,0)$ ⇒ $1+0 → 1$ ⇒ ✓

-ve $b = (2,-2)$ ⇒ $2-4 ⇒ -2$ ⇒ ✓

$\rho = \frac{2}{\|w\|} = \frac{2}{\sqrt{1+4}} = \frac{2}{\sqrt{5}} = 0.894$

Here for $a = (-1,0) ⇒ = \frac{1}{\sqrt{5}}$; $b = (2,-2) ⇒ = \frac{2}{\sqrt{5}}$, $c = (1,0) ⇒ 1/\sqrt{5}$

not a support vector.

⇒ It is not maximizing margin.

→ It's not satisfying all svm conditions.

(d) $2x_1 + 3x_2 = 0.$

⇒ $(w_1, w_2) = (2,3)$

for $a = (-1,0) = 2(-1) + 3(0) = -2$

$b = (2,-2) = 2(2) + 3(-2) = -2$

$c = (1,0) = 2(1) + 0 = 2$

all classified correctly.

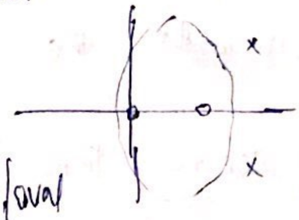
③ $\rho = \frac{2}{\|w\|} = \frac{2}{\sqrt{4+9}} = \frac{2}{\sqrt{13}} < 1$ → ∵ margin should be ≥1. for one-side

⇒ Here margin is not maximizing the width

⇒ So, it's not satisfying svm conditions.

(ii) ⓐ for polynomial kernel, degree = 2

⇒. Decision boundary. It suits for diagram 'D'.



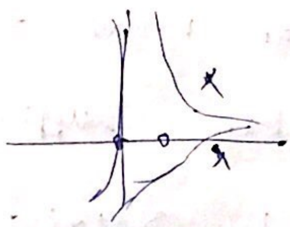
Ⓓ

since degree = 2 ⇒ polynomial (kernel) It's structure will be Elliptical/oval shaped. ⇒ It classifies correctly.

ⓑ polynomial kernel, degree = 3.

$(1+x^Tx)^3$.

⇒. It fits the data



Ⓑ.

ⓒ RBF kernel, $\sigma = 0.5$.

$$K(x,x') = e^{\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)} = e^{(-\gamma \|x-x'\|^2)}$$

$$\therefore \gamma = \frac{1}{2\sigma^2}$$

Diagram.

Here $\sigma = 0.5$ less

⇓

γ = high

It

⇒. diagram Ⓐ is suits well.

$\sigma \propto \frac{1}{\gamma}$

σ = ~~high~~ Low, γ = ~~low~~ High ⇒. SVM depends on just the points that are closest to decision boundary.

γ = low, σ = High ⇒ depends on points that are Farther from it

ⓓ RBF kernel, $\sigma = 1$.

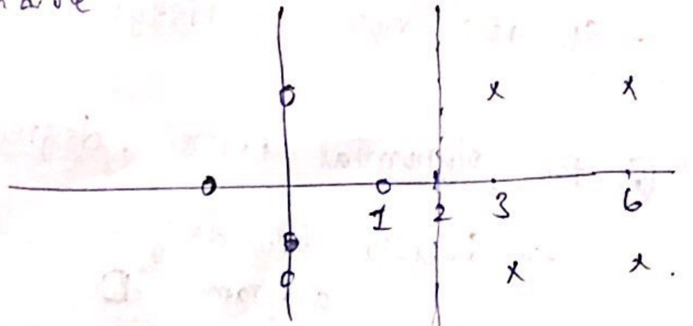
⇓

γ = low.

⇒. diagram Ⓒ decision boundary. a suits wells.

(4) Given data points $+ve = (3,1),(3,-1),(6,1),(6,-1)$
$-ve = (1,0),(0,1),(0,-1),(-1,0)$.

Here it is Linearly seperable data.



So by Linear Kernel we can discriminate the points.

By seeing, this at $(2,0)$ → the Decision boundary can be passing which classifies points correctly & with maximum margin. ⇒ 3 support vectors $(1,0)$, $(3,1)$, $(3,-1)$
a b c.

Linear kernel $= (x^T x)$.

For these support vectors $\alpha \neq 0$. ; For remaining $\alpha = 0$.

$W = \sum \alpha x_i y_i$

⇒ The decision boundary passes through $(2,0)$.
& a, b, c are perpendicular to it.

→ $x_1 + x_2 = 2$
(if $w_1 = w_2$.)

∴ slope $= \frac{3-1}{1-0} = 2$.
→ $m = 2$.

⇒ $1.w_1 + 0.w_2 + b = -1$ → (i)
⇒ $3w_1 + 1w_2 + b = 1$ → (ii)
⇒ $3w_1 + -1w_2 + b = 1$ ⇒ (iii)

by solving them.

$2x + b = -1$
$6x + b = 1$ ← For 'b'

$-4x = -2$
$x = 1/2$.

(or) we can solve using.
$W = \sum \alpha x_i y_i = \sum \alpha K(x, x')$
⇒ $\alpha_1 . (a . \bar{a}) + \alpha_2 (a . \bar{b}) + \alpha_3 (a . c) = -1$.
$\alpha_1 (b . a) + \alpha_2 (b . b) + \alpha_3 (b . c) = 1$
$\alpha_1 (c . a) + \alpha_2 (c . b) + \alpha_3 (c . c) = 1$

⇓
by solving this

add bias within them $a = (1,0)$
⇒ $a = (1\ 0\ 1)$

$\bar{w} = (2.(1/2), 0) = (1,0)$.

$2x + b = -1$ ⇒ ✓ considered as -ve.

So, $b = -2$, $w = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

$2\alpha_1 + 4\alpha_2 + 4\alpha_3 = -1$

$4\alpha_1 + 11\alpha_2 + 9\alpha_3 = +1$

$4\alpha_1 + 9\alpha_2 + 11\alpha_3 = +1$

⇒ after solving this (solved in matlab)

$\alpha_1 = -3.5$
$\alpha_2 = 0.75$
$\alpha_3 = 0.75$

$\Rightarrow W = \sum \alpha\, K(x, x')$

$= -3.5 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + 0.75 \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} + 0.75 \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$

$= \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}$ → w (1, 0), → b (−2)

$\Rightarrow W = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $b = -2$

4.

$\Rightarrow W = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $b = -2$.

5. Limitations of 'support vector machines (svm) on Large dataset :-

→ Generally in svm, ~~the~~ If the dataset has high dimension, it will overfit easily. & if it transform into higher dimensions, then computation is very high & expensive.

→ It is a memory based algorithm initially, i.e, when svm is computing Decision boundary, it needs to stored all data point in memory.

→ It is a Binary class algorithm. If a multi-class problem then we need to build many classifiers ⇒ computation & expensive.

Time-Complexity:-

For multi-class svm.

	OVA	OVO	OVO DAG
Training	$O(L^4 N^3)$	$O(L^2 N^3)$	$O(L^2 N^3)$
classification	$O(L^2 N)$	$O(L^2 N)$	$O(LN)$

→ Approach to solve large dataset problem :-

↳ reduce feature. (dimension) by feature reduction.

↳ parallel svm

→ multi-core system computing.

→ Batch-set training svm. & combine.

→

Algorithm :- For this we can use online-/Incremental svm. (Streaming)

→ Divide the dataset into subsets. & training the models. (svm)

→ get Support vectors. of each svm.

→ Now combine these support vectors & ⇒ It will get good support vectors.

→ It equivalent to training on whole (large) dataset.

(or)

→ we can use SGD (Stochastic Gradient descent) for multi classification & optimization. w.r.t to svm conditions.

It acts as ~~an~~ Incremental- svm., solve large dataset problem.