Reinaldo Daniswara
604840665

Pset 3 M146

1 VC Dimension

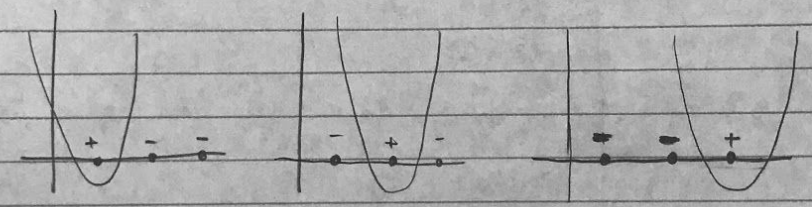
$H = \{ \text{sgn}(ax^2+bx+c) ; a, b, c, \in R \}$

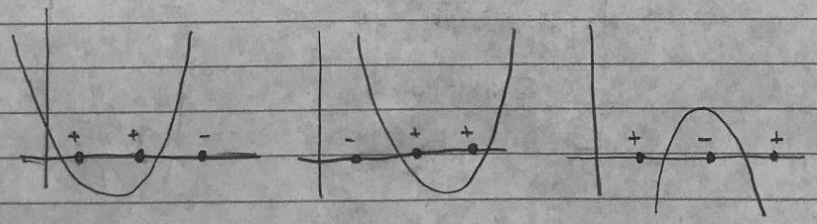
sgn(·) is 1 when · +
0 · −

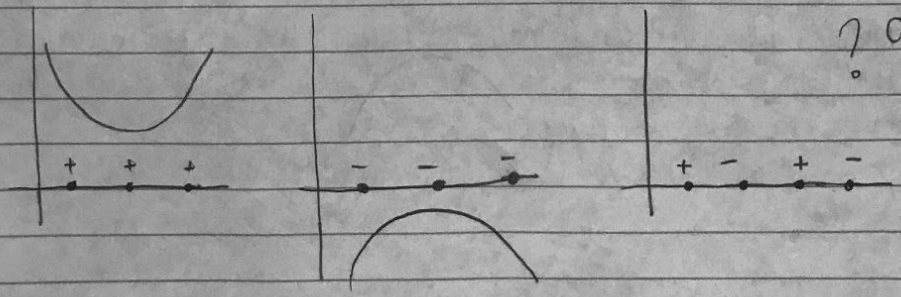
Answer = There is going to be 3 VC dimension

Prove = By contradiction, we will show that quadratic equation could not shatter into alternating 4 VC dimension because quadratic equation can only change sign twice.

Ex:







? Can't be Shattered

2 Kernels

$$K_\beta(x,z) = (1 + \beta x \cdot z)^3$$

$$K_\beta = 1 + 3\beta x_1 z_1 + 3\beta x_2 z_2 + 3\beta^2 x_1^2 z_1^2 + 6\beta^2 x_1 x_2 z_1 z_2 + 3\beta^2 x_2^2 z_2^2 + \beta^3 x_1^3 z_1^3 + 3\beta^3 x_1^2 x_2 z_1^2 z_2 + 3\beta^3 x_1 x_2^2 z_1 z_2^2 + \beta^3 x_2^3 z_2^3$$

Since we know that dot products in high dimensional spaces of

$$K(x,z) = \phi(x)^T \phi(z)$$

We know that it's going to be

$$\phi(\vec{x}) = \begin{bmatrix} 1 \\ \vdots \\ \sqrt{3} x_i \sqrt{\beta} \\ \vdots \\ \sqrt{3}\, x_i^2 \beta \\ \vdots \\ \sqrt{6}\, x_i x_j \beta \\ \vdots \\ x_i^3 \sqrt{\beta^3} \\ \vdots \\ \sqrt{3}\, x_i^2 x_j \sqrt{\beta^3} \\ \vdots \\ \sqrt{3}\, x_i x_j^2 \sqrt{\beta^3} \\ \vdots \end{bmatrix}$$

The differences between $K_\beta(x,z)$ and $K(x,z)$ is $K_\beta(x,z)$ has a parameter β while $K(x,z)$ doesn't have it. The roles of parameter β in $K_\beta(x,z)$ is to scaling the vector. It will scale every single element by a constant β, raise to a certain power.

3 SVM

a. $x_1 = (1,1)^T \quad x_2 = (1,0)^T$

$y_1 = 1 \qquad y_2 = -1$

$w_1 + w_2 \geq 1 \qquad -(w_1) \geq 1$

$1 - w_1 - w_2 \leq 0 \qquad 1 + w_1 \leq 0$

$$L(w,\alpha) = \frac{1}{2}(w_1^2 + w_2^2) + \alpha_1(1 - w_1 - w_2) + \alpha_2(1 + w_1) = z(x).$$

$$d^* = \max_\alpha \min_w L(w,\alpha)$$

minimize in respect to w,

$$\frac{\partial L}{\partial w_1} = w_1 - \alpha_1 + \alpha_2 = 0 \rightarrow w_1 = \alpha_1 - \alpha_2$$

$$\frac{\partial L}{\partial w_2} = w_2 - \alpha_1 = 0 \rightarrow w_2 = \alpha_1 \quad \} \ (1)$$

Sub into $z(x)$.

$$z(x) = \frac{1}{2}\left((\alpha_1 - \alpha_2)^2 + \alpha_1^2\right) + \alpha_1(1 - \alpha_1 + \alpha_2 - \alpha_1) + \alpha_2(1 + \alpha_1 - \alpha_2)$$

$$= \frac{1}{2}\left(2\alpha_1^2 - 2\alpha_1\alpha_2 + \alpha_2^2\right) + (\alpha_1 - \alpha_1^2 + \alpha_1\alpha_2 - \alpha_1^2) + (\alpha_2 + \alpha_1\alpha_2 - \alpha_2^2)$$

$$\frac{\partial z}{\partial \alpha_1} = -2\alpha_1 + \alpha_2 + 1 = 0 \rightarrow \alpha_1 = \frac{\alpha_2 + 1}{2}$$

$$\frac{\partial z}{\partial \alpha_2} = -\alpha_2 + 1 + \alpha_1 = 0 \rightarrow \alpha_2 = \alpha_1 + 1 \quad \}$$

$$\alpha_1 = \frac{\alpha_1 + 1 + 1}{2}$$

$$\frac{\alpha_1}{2} = 1 \qquad \boxed{\alpha_1 = 2 \quad \alpha_2 = 3}$$

Sub $\alpha_1 = 2$ & $\alpha_2 = 3$ to. (1)

$w_1 = 2 - 3 = -1 \qquad w_2 = 2.$

$$\therefore w^* = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

b. With parameter offset b (non zero) we will get

$w_1 + w_2 + b \geq 1$ and $-(w_1 + b) \geq 1$

$1 - w_1 - w_2 - b \leq 0$ $\quad 1 + w_1 + b \leq 0$

Thus,

$$L(w, \alpha, b) = \frac{1}{2}(w_1^2 + w_2^2) + \alpha_1(1 - w_1 - w_2 - b) + \alpha_2(1 + w_1 + b) = z(x)$$

$$\frac{\partial L}{\partial w_1} = w_1 - \alpha_1 + \alpha_2 = 0 \rightarrow w_1 = \alpha_1 - \alpha_2$$

$$\frac{\partial L}{\partial w_2} = w_2 - \alpha_1 = 0 \rightarrow w_2 = \alpha_1 \quad \} \text{①}$$

Sub ① into $z(x)$.

$$z(x) = \frac{1}{2}(\alpha_1^2 - 2\alpha_1\alpha_2 + \alpha_2^2 + \alpha_1^2) + \alpha_1(1 - \alpha_1 + \alpha_2 - \alpha_1 - b) + \alpha_2(1 + \alpha_1 - \alpha_2 + b).$$

$$= \frac{1}{2}(2\alpha_1^2 - 2\alpha_1\alpha_2 + \alpha_2^2) + (\alpha_1 - 2\alpha_1^2 + \alpha_1\alpha_2 - \alpha_1 b) + (\alpha_2 + \alpha_1\alpha_2 - \alpha_2^2 + \alpha_2 b)$$

$$\frac{\partial z}{\partial \alpha_1} = 2\alpha_1 - \alpha_2 + 1 - 4\alpha_1 + \alpha_2 - b + \alpha_2 = 0$$

$$= -2\alpha_1 + \alpha_2 + 1 - b = 0 \rightarrow -2\alpha_1 = -\alpha_2 - 1 + b$$

$$\alpha_1 = \frac{\alpha_2 + 1 - b}{2}$$

$$\frac{\partial z}{\partial \alpha_2} = -\alpha_2 + \alpha_1 + 1 + b = 0 \rightarrow \alpha_2 = \alpha_1 + 1 + b$$

$$\alpha_1 = \frac{\alpha_1 + 1 + b + 1 - b}{2} \qquad \frac{\alpha_1}{2} = 1$$

$\alpha_1 = 2 = \alpha_2$

thus, $b = -1$

sub to ①

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

thus, $\rightarrow 2 = \frac{1}{2}$

without offset $w^* = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ $\quad 2 = \frac{1}{\sqrt{5}}$

By having offset, of course w^* value is changed, and we will have a larger margin.

4.1 Feature Extraction

A. Implemented in the code

B. Implemented in the code

C. Implemented in the code

D. Dimensionality of the all the tweet is (630, 1140930). Thus, splitting into training and test data will result in dimensionality of training data (560, 1014160) and dimensionality of test data (70, 126770)

4.2 Hyper-parameter Selection for a Linear-Kernel SVM

A. Implemented in the code

B. It is beneficial to maintain class proportions roughly the same across folds because if the proportions are not balance or not roughly the same across folds there will be an inaccuracy of measurement due to one class is more represented than the other. For example, if one class is not represented well enough, it is going to be hard to determine the decision boundary for that particular class because it does not have enough data to learn.

C. Implemented in the code

C	accuracy	F1-score	AUROC
10^{-3}	0.7089	0.8297	0.8105
10^{-2}	0.7107	0.8306	0.8111
10^{-1}	0.8060	0.8755	0.8576
10^{0}	0.8146	0.8749	0.8712
10^{1}	0.8182	0.8766	0.8696
10^{2}	0.8182	0.8766	0.8696
Best C	100	100	100

In svm, parameter C uses to determine the hyperplane or boundary between data that we want to classify. It is clear that, larger the value of C, it will choose a small value of margin to avoid any misclassified points. Same instance happen with the small value of C, it will choose a larger value of margin, which tend to have more misclassified points. According to those ideas, for any method of performance measurer (accuracy, F1-score, and AUROC) we will see that larger the value of C, the performance will also get better. I also did an experiment to try to have larger values of C, and it seems that the performance could not get any better for all of the performance measurer method. We may need to try another method to get a closer value to 1.

4.3 Test Set performance

A. Best C is 100

B. implemented in the code

C.

Metric	Linear SVM score
Accuracy	0.7429
F1_score	0.4375
AUROC	0.7464