CSM146: Homework 3

Alex Gilbert
904-770-190
Disc. 1D

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

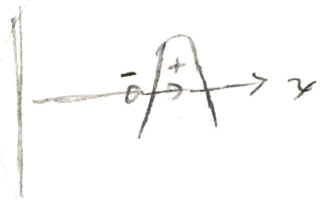
↳ 1 when positive
0 o/w

$\boxed{\text{VC Dimension} = 3}$

Proof...

VC ≥ 1 :

VC ≥ 2 :



VC ≥ 3 : * all 3 same sign obvious extension of above cases

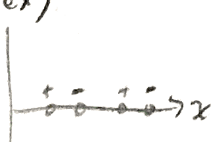
* If 2 of same sign adjacent, simply place 3rd point @ x less than or greater than zero of parabola from corresponding example above

Final case : same sign points not adjacent...

Note, if choice of labels switched, simply flip parabola

VC < 4 : Consider 4 points arbitrarily placed @ points along the x axis ∴ sign switches 3 times btw/n points

ex)



such that adjacent points have alternating signs - note sign given to each arbitrary. If 2 adjacent points have alternating signs, they must be separated by zeros of the parabola. ∃ ≤ 2 zeros of the parabola, ∴ for any placement of the 4 points ∄ enough zeros to separate the 3 sign switches. Therefore the given classifier can't shatter 4 points => $\boxed{\text{VC Dimension} = 3}$

CSM146: Homework 3

Alex Gilbert
904-770-190
Disc. 10

2.) $K_\beta(x,z) = (1+\beta x\cdot z)^3 \mid \beta > 0;\ x, z \in \mathbb{R}^2$. Find $\phi_\beta(\cdot)$.

$\hookrightarrow K_\beta(x,z) = \phi_\beta(x)^T \phi_\beta(z)$

$$(1+\beta x\cdot z)^3 = (1+\beta x^T z)^3 = (1+\beta x_1 z_1 + \beta x_2 z_2)^3$$

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

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

found using Wolfram Alpha

$$= \left(\begin{bmatrix} 1 \\ \sqrt{3\beta}\, x_1 \\ \sqrt{3\beta}\, x_2 \\ \sqrt{6}\beta\, x_1 x_2 \\ \sqrt{3}\beta\, x_1^2 \\ \sqrt{3}\beta\, x_2^2 \\ \sqrt{3}\beta^{3/2} x_1^2 x_2 \\ \sqrt{3}\beta^{3/2} x_1 x_2^2 \\ \beta^{3/2} x_1^3 \\ \beta^{3/2} x_2^3 \end{bmatrix}\right)^T \left(\begin{bmatrix} 1 \\ \sqrt{3\beta}\, z_1 \\ \sqrt{3\beta}\, z_2 \\ \sqrt{6}\beta\, z_1 z_2 \\ \sqrt{3}\beta\, z_1^2 \\ \sqrt{3}\beta\, z_2^2 \\ \sqrt{3}\beta^{3/2} z_1^2 z_2 \\ \sqrt{3}\beta^{3/2} z_1 z_2^2 \\ \beta^{3/2} z_1^3 \\ \beta^{3/2} z_2^3 \end{bmatrix}\right)$$

$$\therefore\ \boxed{\phi_\beta(x) = \begin{bmatrix} 1 \\ \sqrt{3\beta}\, x_1 \\ \sqrt{3\beta}\, x_2 \\ \sqrt{6}\beta\, x_1 x_2 \\ \sqrt{3}\beta\, x_1^2 \\ \sqrt{3}\beta\, x_2^2 \\ \sqrt{3}\beta^{3/2} x_1^2 x_2 \\ \sqrt{3}\beta^{3/2} x_1 x_2^2 \\ \beta^{3/2} x_1^3 \\ \beta^{3/2} x_2^3 \end{bmatrix}}$$

β adds a constant multiplier equal to β^($\frac{1}{2}$ degree of the term) to each term of the feature vector & the kernel function of $K(x,z)$.

This allows you to weight the importance of higher order terms, in direct or indirect proportion to their order.

CSM146: Homework 3

Alex Gilbert
904-770-190
Disc. 1D

3.) a. $y w^T x = 1$

$1 \Rightarrow \quad 1(w_1, w_2) \cdot (1,1) = 1$

$w_1 + w_2 = 1$

$2 \Rightarrow \quad -1(w_1, w_2) \cdot (1,0) = 1$

$w_1 = -1 \rightarrow w_2 = 2 \quad \therefore \quad \boxed{w^* = \begin{bmatrix} -1 \\ 2 \end{bmatrix}}$

b. $y(w^T x + b) = 1$

$1 \Rightarrow \quad 1\left[(w_1, w_2)(1,1) + b\right] = 1$

$w_1 + w_2 + b = 1$

$2 \Rightarrow \quad -1\left[(w_1, w_2)(1,0) + b\right] = 1$

$-w_1 - b = 1$

$w_2 = 2$

minimize $\|w\|$ $\therefore$ $w_1 = 0$

$\therefore b = 1$

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

Note $\| w^*_{b=0} \| = 5 < \| w^*_{b \neq 0} \| = 4$

$\therefore$ allowing offset yields
larger margin

4.1.c) Feature Extraction and Train/Test Splits

I have completed the feature extraction and generated the train/test splits.

4.2.b) Maintaining Class Proportions in K-Fold Cross-Validation

It is important to maintain the proportions of each class across folds in K-fold validation—called stratification—because ideally, one wants each fold to be representative of the entire dataset. This is achieved by distributing the class instances to each fold proportionally to how they occur in the entire sample dataset…which hopefully is a good model for the underlying distribution.

4.2.d) Selecting the Optimal Hyperparameter

The table to the right shows the results of performing K-fold stratified cross validation on the parameter C of linear-kernel SVM classifiers, for three different classifiers—accuracy, F1-score, and area under the receiver operating characteristic curve (AUROC). We see that for judging by accuracy, the optimal C = 10 (Accuracy = 0.8300); by F1-score, the optimal C = 10 (F1-Score = 0.8839); by AUROC, the optimal C = 1 (AUROC = 0.8994). Note that for many of all three of the metrics, multiple C's resulted in the same performance score. Accordingly, we chose the smallest such C with the max performance value, to get the largest margin. Typically, it is impossible to say which will improve test results the most, as this depends on the data you encounter during testing, but I found that any of the matching C's actually yielded identical test performance.

C	accuracy	F1-score	AUROC
10^{-3}	0.7102	0.8306	0.8542
10^{-2}	0.7246	0.8370	0.8542
10^{-1}	0.8121	0.8782	0.8809
10^{0}	0.8211	0.8779	0.8994
10^{1}	0.8300	0.8839	0.8994
10^{2}	0.8300	0.8839	0.8994
best C			

4.3.c) Test Set Performance

Performance Metric Used	Value of C parameter	Test Performance Score
Accuracy	10	0.7391
F1_Score	10	0.4706
AUROC	1	0.7431