Cody Hubbard 004843389
CSM146, Winter 2018
Problem Set 3

Problem 1

Solution: The VC dimension of $H = sgn(ax2 + bx + c); a, b, c, \in R$ is 3. First consider the shape of our H, it is the set of parabolas scaled using a, b, c. Due to the single dimensionality of our x values, for any three points we can find a set of scalars that allow us to shatter those points without much difficulty so VC dim ≥ 3. However consider 4 points, say $x = 1, 2, 3, 4$ with corresponding labels $+, -, +-$Due to the nature of a parabola, and our alternating labels, no set of scalars a, b, c will allow us to shatter these points. So it is impossible for our parabola it to obtain a third critical point, changing the functions direction again to satisfy the 4th (or 1st) point's label. Lastly due to the single-dimensionality of our x this case can cover all 4-point sets of x. No matter how spread out or close together the points are this will still hold. Thus this shows that for all 4-point sets of x there exits a set of labels $(+, -+, -)$ that are not shatterable by H, proving the VC-dimension< 4. So if the VC-dim ≥ 3 and < 4, VC-dim =3 as to be shown.

Problem 2

Solution: $K_\beta(x, z) = \beta^3(x^3 \cdot z^3) + 3\beta^2(x^2 \cdot z^2) + 3\beta(x \cdot z) + 1$ The corresponding feature map is $\phi_\beta(x) = [\beta^{\frac{3}{2}}x^3, \beta\sqrt{3}x^2, \sqrt{3\beta}x, 1]$. That is it is the third degree polynomial expansion of the original feature scaled by some β. The similarities/differences to the other kernel, $K(x, z) = (1 + x \cdot z)^3$ is the non scaling. that is the corresponding feature map to this kernel is still the third degree polynomial expansion but is not scaled by β. The role of β is to scale the distance, or level of similarity, calculated by the dot product.

Problem 3

(a) Solution: minimize $\frac{1}{2}\left\|w\right\|^2$
subject to $y_n w^t x_n \geq 1$ where $n = 1, ..., N$
given $x_1 = [1, 1]^t$ $y_1 = 1$ and $x_2 = [1, 0]^t$ $y_2 = -1$
we can form a system of equations to find the minimum values of w

$$(1)(w^t x_1) \geq 1 \rightarrow (1)[\ w_1 \quad w_2\ ]\begin{bmatrix} 1 \\ 1 \end{bmatrix} \geq 1 \rightarrow (w_1 + w_2) \geq 1$$

$$(-1)(w^t x_2) \geq 1 \rightarrow (-1)[\ w_1 \quad w_2\ ]\begin{bmatrix} 1 \\ 0 \end{bmatrix} \geq 1 \rightarrow (-w_1) \geq 1$$

with this system of equations we can see that $w_1 \leq -1$ and $w_2 \geq 2$ thus the values of w_1 and w_2 that minimize $\frac{1}{2}\left\|w\right\|^2$ are $w_1 = -1$ and $w_2 = 2$ that is $w* = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$.

(b) Solution: Allowing the offset parameter b to be nonzero changes the classifier to be horizontal line. The new $w*$ is $w* = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $b* = -\frac{1}{2}$. The margin stays the same (.5) but it no longer satisfies our constraints, otherwise this $w*$would give us a better minimum.

Problem 4.1

(a) Solution: -Implemented extract_dictionary()

(b) Solution: -Implemented extract_feature_vectors()

(c) Solution: -Split the feature matrix and label array

(d) Solution: (630, 1811)

Problem 4.2

(a) Solution: -Implemented performance()

(b) Solution: -Implemented cv_performance()

(c) Solution: -Implemented select_param_linear()
-Used StratifiedKFold() to split the data for 5-fold CV
It is beneficial to maintain class proportions across folds because we want our folds to have consistency. If say one or two folds had all positive or all negative examples it would skew that folds results.

(d) Solution:

C	accuracy	F1-score	AUROC
0.001	0.7089	0.8296	0.5
0.01	0.7107	0.8305	0.5031
0.1	0.8060	0.8754	0.7187
1.0	0.8146	0.8748	0.7531
10.0	0.8181	0.8765	0.7591
100.0	0.8181	0.8765	0.7591
best C	10.0	10.0	10.0

As visible in the above table regardless of the performance metric used the higher the value of C the more accurate our model. However the increase from 10 to 100 had such little return in accuracy increase that it is not even detectable within 17 decimal places.

Problem 4.3

(a) Solution: -Trained a linear-kernel SVM with $C = 10$

(b) Solution: -Implemented performance_test()

(c) Solution:

C=10	
Metric	Test Data Performance
accuracy,	0.742857142857
F1-score	0.4375
AUROC	0.625850340136