CM146, Fall 2017
Problem Set 3: Computational Learning Theory, Kernel, SVM
Due Feb 27, 2018

1 Problem 1

Solution: VC dimension of H is 3. First let's prove $VC(H) < 4$. Pick points $x = 0, 1, 2, 3$ and assign values $1, 0, 1, 0$, respectively. Assume the contrary, *i.e.* there exists $f(x) = ax^2 + bx + c$ such that $f(0) > 0$, $f(1) < 0$, $f(2) > 0$, $f(3) < 0$. A second degree polynomial has at most 2 roots. Since f must be both negative and positive, and it is continuous, it must have 2 distinct roots. Call these x_1 and x_2. Walking along the x axis, we can switch from negative to positive values of f when we go through one of these roots. We know we can only switch 2 times, but our setup of $1, 0, 1, 0$ requires at least 3 switches which gives a contradiction. So, $VC(H) < 4$.

Claim: H can always shatter a set of three points. Proof: Borrowing the notation, let's look at how the three points, looking from $-\infty$ to $+\infty$, are mapped. For any mapping, we will have to change the sign of f at most twice (when we have $(1, 0, 1)$ and $(0, 1, 0)$). Since we can have 2 roots, we know we can always shatter a set of 3 points. So, $VC(H) \geq 3$.

So, $VC(H) = 3$.

2 Problem 2

Solution:

$$
\begin{aligned}
K_\beta(x,y) = (1+\beta x\cdot y)^3 &= \beta^3(x\cdot y)^3 + 3\beta^2(x\cdot y)^2 + 3\beta x\cdot y + 1 \\
&= \beta^3 x_2^3 y_2^3 + 3\beta^3 x_1 x_2^2 y_1 y_2^2 \\
&+ 3\beta^3 x_1^2 x_2 y_1^2 y_2 + \beta^3 x_1^3 y_1^3 \\
&+ 3\beta^2 x_2^2 y_2^2 + 6\beta^2 x_1 x_2 y_1 y_2 \\
&+ 3\beta^2 x_1^2 y_1^2 + 3\beta x_2 y_2 + 3\beta x_1 y_1 + 1
\end{aligned}
$$

So,

$\phi_\beta(x) =$
$\begin{bmatrix} 1 & \sqrt{3\beta}x_1 & \sqrt{3\beta}x_2 & \sqrt{3}\beta x_1^2 & \sqrt{3}\beta x_2^2 & \sqrt{6}\beta x_1 x_2 & \beta^{3/2}x_1^3 & \beta^{3/2}x_2^3 & \sqrt{3}\beta^{3/2}x_1^2 x_2 & \sqrt{3}\beta^{3/2}x_1 x_2^2 \end{bmatrix}^T.$

Also,

$\phi(x) = \begin{bmatrix} 1 & \sqrt{3}x_1 & \sqrt{3}x_2 & \sqrt{3}x_1^2 & \sqrt{3}x_2^2 & \sqrt{6}x_1 x_2 & x_1^3 & x_2^3 & \sqrt{3}x_1^2 x_2 & \sqrt{3}x_1 x_2^2 \end{bmatrix}^T.$

So, notice that $\phi(x) = \phi_\beta(x\beta^{-1/2})$ for any $\beta > 0$. Therefore, β can be used as a scaling factor, that inflates/deflates distance.

3 Problem 3

(a) **Solution:** Maximize the margin to get the best margin:

$$\begin{aligned}
&\max\left\{\min\left[\sqrt{2}\sin\left(\frac{\pi}{4}-\theta\right), \sin\theta\right]\right\} \\
=&\max\left\{\min\left[\cos\theta-\sin\theta, \sin\theta\right]\right\}, \quad 0<\theta<\frac{\pi}{4} \\
=&\frac{\sqrt{5}}{5}, \quad \theta=\tan^{-1}\left(\frac{1}{2}\right).
\end{aligned}$$

So, $\omega^* = \begin{bmatrix}-1 & 2\end{bmatrix}^T$.

(b) **Solution:** The plane that optimally separates the two points is $y=\frac{1}{2}$. The margin is $\frac{1}{2}$ and $\omega^* = \begin{bmatrix}0 & 2\end{bmatrix}^T$, $b^* = -1$.

4 Problem 4

4.1 Feature Extraction

Solution: (630, 1811)

4.2 Hyper-parameter Selection for a Linear-Kernel SVM

(a)

(b)

(c) **Solution:** It is important to maintain the proper ratio of classes across fields because improper separation of data can make a model better of worst depending on how incorrectly skewed the training (and therefore the testing) data is compared to the actual ratio. This is not the point of course. We want to see which model is better for the data, not which data is better for the model.

(d) **Solution:** Optimal solutions for accuracy, F1-score, and AUROC are 10, 10, and 10 respectively.

C	accuracy	F1-score	AUROC
10^{-3}	0.7089	0.8297	0.5000
10^{-2}	0.7107	0.8306	0.5031
10^{-1}	0.8060	0.8749	0.7188
1	0.8146	0.8749	0.7531
10	0.8182	0.8766	0.7592
100	0.8182	0.8766	0.7592

We can notice that the models tend to significantly increase in accuracy (especially the AUROC score) as we increase C (that is, as we increase the importance of model simplicity vs. training data fit). This indicates that the model by itself would overfit the data. We see little gain in performance as C is further increased.

4.3 Test Set Performance

Solution: For the accuracy metric, the test performance is 0.7429.

For the F1-score metric, the test performance is 0.4375.

For the AUROC metric, the test performance is 0.6259.