# CS146 HW3

## 1

The VC dimension is 3.
If we choose 3 points $x_1, x_2, x_3$, for any label $y_1, y_2, y_3$ assigned to the points. We can just let $ax_i^2 + bx_i + c = 1$ if $y_i = 1$ or = -1 if $y_i = 0$. Then we get 3 equations regard of a,b,c. We can definitely find values of a,b,c that satisfies the requirement. So H can shatter 3 points, VC(H)$\geq$ 3.
For any 4 points $x_1, x_2, x_3, x_4$, that $x_1 \leq x_2 \leq x_3 \leq x_4$, if the label is $y_1 = 1, y_2 = 0, y_3 = 1, y_4 = 0$. Then there is no a,b,c that can correctly predict those. So VC(H)¡4, which shows that VC(H)=3.

## 2

Because x*z = z*x, $K_\beta(x, z) = (1 + \beta x * z)^3 = (1 + \beta z * x)^3 = K_\beta(z, x)$.
$K_\beta(x, z) = \beta^3(x \cdot z)^3 + 3\beta^2(x \cdot z)^2 + 3\beta x \cdot z + 1 = \beta^3(x_1z_1 + x_2z_2)^3 + 3\beta^2(x_1z_1 + x_2z_2)^2 + 3\beta(x_1z_1 + x_2z_2) + 1 = \beta^3((x_1z_1)^3 + 3(x_1z_1)^2x_2z_2 + 3x_1z_1(x_2z_2)^2 + (x_2z_2)^3) + 3\beta^2((x_1z_1)^2 + 2x_1z_1x_2z_2 + (x_2z_2)^2) + 3\beta(x_1z_1 + x_2z_2) + 1$
$\Theta(x) = (1, \sqrt{3\beta}x_1, \sqrt{3\beta}x_2, \sqrt{3}\beta x_1^2, \sqrt{3}\beta x_2^2, \sqrt{6}\beta x_1x_2, \sqrt{3\beta^3}x_1^2x_2, \sqrt{3\beta^3}x_1x_2^2, \sqrt{\beta^3}x_1^3, \sqrt{\beta^3}x_2^3)$.
$\beta$ is a parameter that can be used to scale the feature map up and down.

## 3

(a)
The maximum-margin linear classifier should be equally distant from $x_1$ and $x_2$. So the classifier passes the middle point between $x_1$ and $x_2$, which is $(1, 0.5)^T$. $\omega^* = (\omega_1, -2 * \omega_1)$. $y_1\omega^Tx_1 = -\omega_1 \geq 1$ , $y_2\omega^Tx_2 = -omega_1 \geq 1$. We get $\omega_1 \leq -1$, so the smallest $\|\omega\|$ is when $\omega_1 = -1$, $\omega^* = (-1, 2)$.

(b)
Let the classifier be $(\omega_1, \omega_2)$, since the classifier should still pass the middle point (1,0.5), $\omega_1 + 0.5\omega_2 + b^* = 0$. Then we get $\omega_2 = -2b^* - 2\omega_1$. Plug thisinto

$y_n\omega^T x_n \geq 1$, we get $\omega_1 - 2b^* - 2\omega_1 + b^* = -\omega_1 - b \geq 1$, $\omega_1 + b \leq -1$. So $b \geq -1 - \omega_1$. $\|\omega^*\|^2 = \omega_1^2 + \omega_2^2 = 5\omega_1^2 + 8\omega_1 + 4b^2$. This is minimized when $\omega_1 = 0$, so $b \geq -1 \rightarrow b = -1$. So $\omega^* = (0, 2)$ and $b^* = -1$. Compared with the solution without offset, $\omega_2$ is the same but $\omega_1$ is different.

# 4

## 4.1

(d) I have finished the extraction and generated the splits.

## 4.2

(d)

| 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 | 10.0 | 10.0 | 1.0 |

For all three metrics, the score tends to increase when C is small and increasing. But after C reaches a certain value, 10.0 for accuracy and F1-score, 1.0 for AUROC, the score starts decreases or not increase.

## 4.3

(a) I choose C=10.0 because this is the setting with overall best score.
(c) For C=10.0
The performance employed accuracy metric is 0.7428571428571429.
The performance employed f1-score metric is 0.43749999999999994.
The performance employed AUROC metric is 0.7463556851311952.