hw3

Xiwei Ma

November 2018

1 VC Dimension

The VC dimension of H should be 3.
First, we can prove $VC(H) >= 3$ by showing there exists 3 points can be shattered. Let $sgn(ax_i^2 + bx_i + c) = 1$ for positive label $y_i = 1$ and $sgn(ax_i^2 + bx_i + c) = 0$ for negative label $y_i = -1$. Then we can get 3 equations that can always be satisfied by setting 3 variables a,b,c.
Then we can prove $VC(H) < 4$ by showing no 4 points can be shattered. Suppose there are 4 points $x_1 < x_2 < x_3 < x_4$, labels y_1 and y_3 are positive while the other two labels are negative. Then there does not exist a $sgn(ax^2 + bx + c)$ equation can split them.
Combine $VC(H) >= 3$ and $VC(H) < 4$, $VC(H) = 3$.

2 Kernels

$K_\beta(x, z) = (1 + \beta x \cdot z)^3 = (1 + \beta(x_1z_1 + x_2z_2))^3$
$= \beta^3(x_1^3z_1^3 + 3x_1^2z_1^2x_2z_2 + 3x_1z_1x_2^2z_2^2 + x_2^3z_2^3) + 3\beta^2(x_1^2z_1^2 + 2x_1z_1x_2z_2 + x_2^2z_2^2) + 3\beta(x_1z_1 + x_2z_2) + 1$
Thus,
$\phi_\beta(x) = (1, \sqrt{3\beta}x_1, \sqrt{3\beta}x_2, \sqrt{3}\beta x_1^2, \sqrt{6}\beta x_1x_2, \sqrt{3}\beta x_2^2, \sqrt{\beta^3}x_1^3, \sqrt{3\beta^3}x_1^2x_2, \sqrt{3\beta^3}x_1x_2^2, \sqrt{\beta^3}x_2^3)^T$

When $K(x, z) = (1 + x \cdot z)^3$ where $\beta = 1$, the coefficients of terms would be $\beta^{1/2} = \beta^{3/2} = \beta$, so we can get $K(x, z) = (1 + x \cdot z)^3 = K_\beta(x, z)$. The parameter β is used to scale map up and down, such that when $0 < \beta < 1$, the lower-order terms would have more weight while higher-order terms have less weight. And it is the reverse when $\beta > 1$. When β approaches ∞, K_β would only get constant and low-order terms remain.

3 SVM

(a)
To make both support vectors $y_1w^Tx_1 = 1$ and $y_2w^Tx_2 = 1$, we can compute to get

$$w^* = [-1, 2]^T$$

.

(b)
Let the classifier be $[w1, w2]^T$. In this case, the previous equations would become $y_1 w^T x_1 + b = 1$ and $y_2 w^T x_2 + b = 1$. So $w1 + w2 + b = 1$ and $w1 + b = -1$, then we get $w_2 = 2$. And to minimize $|w^*|$, $w_1 = 0$, and then $b = -1$. Thus,

$$w^* = [0, 2]^T, b^* = -1$$

4 Twitter analysis using SVMs

4.1 Feature Extraction

(a) Implemented.
(b) Implemented.
(c) Implemented.
(d) Finished the extraction and split requirements.
train_X = X[:560,:] train_y = y[:560] test_X = X[560:,:] test_y = y[560:]

4.2 Hyper-parameter Selection for a Linear-Kernel SVM

(a) Implemented.
(b) Implemented. Because as we learned in class, the distribution of data should be similar in training data and test set. We make the assumption that the training set should be representative. Thus, to make sure the assumption hold, we should maintain the class proportions across folds during training.
(c) Implemented.
(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	$10,10^2$	$10,10^2$	10^0

For accuracy and F1-score, the performance score increases as c increases until C is large enough, then the performance score stays the same. For AUROC matrix, the score first increases and then decrease a little bit to a constant as c increasing.

4.3 Test Set Performance

(a) Implemented.
(b) Implemented.
(c) The results are:

accuracy, c = 10: 0.7428571428571429
F1_score, c = 10: 0.43749999999999994
AUROC, c = 1: 0.7463556851311952