# CM146, Fall 2018
# Problem Set 03: Jonathan Chu

November 18, 2018

## 1 VC Dimension

(a) The VC Dimension of H is 3. To prove this we will show that VC $\geq$ 3 and VC $<$ 4.

The model $ax^2 + bx + c; a, b, c, \in R$ can change sign twice at any values of x, beginning with either sign {-1, 1} at $x = -\infty$. Therefore any label assignment to any three points can be separated by the model, and VC $\geq$ 3.

With four points, however, our model will not be able to separate the data in every case. Consider the case of $x_1 \leq x_2 \leq x_3 \leq x_4$ (all spatial configurations of four examples must satisfy this property) with $x_1 = x_3 = -1$ and $x_2 = x_4 = 1$. Since there are three sign changes as x goes from $-\infty$ to $\infty$, our hypothesis space does not contain a model that can separate these points.

## 2 Kernels

(a) Expanding the kernel,

$K_\beta(\mathbf{x}, \mathbf{z}) = 1 + 3(\beta\mathbf{x}\cdot\mathbf{z})^2 + 3(\beta\mathbf{x}\cdot\mathbf{z}) + (\beta\mathbf{x}\cdot\mathbf{z})^3 =$
$1 + 3\beta^2(x_1^2z_1^2 + 2x_1z_1x_2z_2 + x_2^2z_2^2) + 3\beta(x_1z_1 + x_2z_2) + \beta^3(x_1^3z_1^3 + 3x_1z_1x_2^2z_2^2 + 3x_1^2z_1^2x_2z_2 + x_2^3z_2^3)$

$\phi_\beta(\mathbf{x})^T\phi_\beta(\mathbf{z}) = K_\beta(\mathbf{x}, \mathbf{z})$
$\Rightarrow$ for $\mathbf{y} \in \mathbb{R}^2$,
$$\phi_\beta(\mathbf{y}) = (1, \sqrt{3}\beta y_1^2, \sqrt{3}\beta y_1y_2, \sqrt{3}\beta y_2^2, \sqrt{3\beta}y_1, \sqrt{3\beta}y_2,$$
$$\sqrt{\beta^3}y_1^3, \sqrt{3\beta^3}y_1y_2^3, \sqrt{3\beta^3}y_1^2y_2, \sqrt{\beta^3}y_2^3)$$

$K(\mathbf{x}, \mathbf{z}) = (1 + \mathbf{x} \cdot \mathbf{z})^3$ is equivalent to $K_\beta(\mathbf{x}, \mathbf{z}) = (1 + \beta \mathbf{x} \cdot \mathbf{z})^3$ with $\beta = 1$. The parameter $\beta$ acts as a coefficient for elements of the feature mapping, with high $\beta$ placing more weight on higher degree elements. It is an additional parameter that gives even more flexibility in the feature map.

# 3 SVM

(a) By graphing the data, it is clear that the line separating the data with maximum margin is one with slope $\frac{1}{2}$, passing through point $(1, \frac{1}{2})$. In other words, $\frac{w_1^*}{w_2^*} = -\frac{1}{2}$.

The two constraints we must satisfy are:

$n = 1 : w_1^* + w_2^* \geq 1$

$n = 2 : -w_1^* \geq 1$

By inspection, $w_1^* = -1, w_2^* = 2$ satisfy both constraints as equalities and minimize $\|w^*\|$.

(b) With the additional parameter b, we seek a weight vector $w^*$ with magnitude less than $\sqrt{5}$, the magnitude from part (a).

Geometrically, it is obvious that the line maximizing the margin $\gamma$ is a horizontal line through the point $(1, \frac{1}{2})$

$\Rightarrow w_1^* = 0, w_2^* > 0$

The new constraints we must satisfy are:

$n = 1 : w_2^* + b \geq 1$

$n = 2 : -b \geq 1$

$\Rightarrow b = -1$

$\Rightarrow w_2^* = 2$

The magnitude $\|w^*\| = 2$

# 4 Twitter analysis using SVMs

## 4.1 Feature Extraction

Done.

### 4.2 Hyper-parameter Selection for a Linear-Kernel SVM

It's beneficial to maintain class proportions across folds because a fold without any regulated proportion could be less representative of the actual data. In extreme cases, a train or test set in a particular fold could be missing examples of a certain label in which case the training and test error values will be far off from reality.

For example, a training set containing no positively labeled examples would simply predict negative always and achieve 0 training error, but it would perform poorly on the test set, where all the positive examples have been placed.

| 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 | 0.8755 | 0.7188 |
| $10^0$ | 0.8146 | 0.8749 | 0.7531 |
| $10^1$ | 0.8182 | 0.8766 | 0.7592 |
| $10^2$ | 0.8182 | 0.8766 | 0.7592 |
| best C | $10^2$ | $10^2$ | $10^2$ |

The score seems to increase as C increases, for every metric. With every metric, the value of C with the best score was $10^2$.

### 4.3 Test-Set Performance

With $C = 10^2$,

| Metric | Test Performance Score |
|---|---|
| Accuracy | 0.7429 |
| F1-Score | 0.4375 |
| AUROC | 0.6259 |