Intro To Machine Learning

Bashar Beshoti (207370248) , Ayal Kaabia (322784760)

July 9, 2024

Course Information

- **Course Title:** Intro To Machine Learning
- **Course Code:** 203.4770
- **Assignment :** 3
- **Due Date :** 09.07.2024

PAC, VC dimension, Bias vs Variance

Section 1: VC Dimension of a Circle Classifier

A circle (r, c) is defined by its center c and its radius r. Consider the following family of classifiers:

$$H = \{h_{r,c} : r \in \mathbb{R}, c \in \mathbb{R}^2\}$$

where $h_{r,c}(x) = 1$ if x is inside the circle (r, c) and $h_{r,c}(x) = 0$ otherwise.

We aim to find the VC dimension of this class with a full proof.

Solution

The function $h_{r,c}(x)$ is defined as:

$$h_{r,c}(x) = \begin{cases} 1 & \text{if } x \text{ is inside the circle } (r, c) \\ 0 & \text{otherwise} \end{cases}$$

Let's analyze the shattering of different sample sets: 1. **For 1 sample (2^1 possible classifications)**: - The sample is either inside or outside the circle, leading to 2 possible classifications.

2. **For 2 samples (2^2 possible classifications)**: - Both samples are outside the circle. - One sample is inside the circle, and the other is outside. - Both samples are inside the circle.

This results in 4 possible classifications, which are all realizable by circles.

3. **For 3 samples (2^3 possible classifications)**: - All samples are outside the circle. - One sample is inside the circle, and two are outside. - Two samples are inside the circle, and one is outside. - All samples are inside the circle.

This results in 8 possible classifications, which are all realizable by circles.

4. **For 4 samples (2^4 possible classifications)**: - We cannot shatter all possible classifications. For instance:

$$\begin{pmatrix} - & + \\ + & - \end{pmatrix}$$

It is not possible to classify the samples as shown above since it would require the circle to include some positive samples while excluding others that are closer. Therefore, there exists at least one set of 4 points that cannot be shattered by circles.

Thus, the VC dimension of this class of circles is:

$$\boxed{\text{VC-Dim} = 3}$$

Section 2: Sample Complexity for a Depth-1 Decision Tree

Consider a training set $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ where $x_i \in \{0,1\}^3$. Each sample has 3 Boolean features $\{X_1, X_2, X_3\}$. The classification rule is:

$$Y = (X_1 \vee X_2) \wedge \neg(X_1 \vee \neg X_2)$$

We attempt to learn the function $f : X \to Y$ using a "depth-1 decision tree". A "depth-1 decision tree" is a tree with a root and two leaves, each at distance 1 from the root.

Solution

Given the depth-1 decision tree, the function can be learned with a VC dimension $d = 2$.

Theorem (The Fundamental Theorem of Statistical Learning)

Let $\mathcal{H}$ be a hypothesis class of binary classifiers. Then, there are absolute constants C_1, C_2 such that the sample complexity of PAC learning $\mathcal{H}$ is

$$C_1 \frac{d + \log(1/\delta)}{\epsilon} \leq m_{\mathcal{H}}(\epsilon, \delta) \leq C_2 \frac{d \log(1/\epsilon) + \log(1/\delta)}{\epsilon}$$

Furthermore, this sample complexity is achieved by the ERM learning rule.

Using the PAC learning framework, the sample complexity for learning with VC dimension d and error ϵ with confidence $1 - \delta$ is given by:

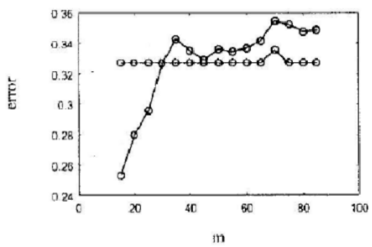
$$\frac{2 + \log\left(\frac{1}{\delta}\right)}{\epsilon} \leq m_H(\epsilon, \delta) \leq \frac{2 \log\left(\frac{1}{\epsilon}\right) + \log\left(\frac{1}{\delta}\right)}{\epsilon}$$

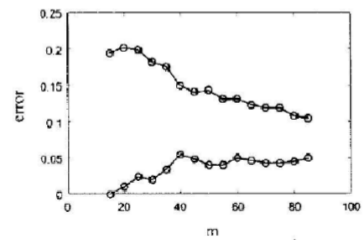
Therefore, the sample complexity for this learning problem is:

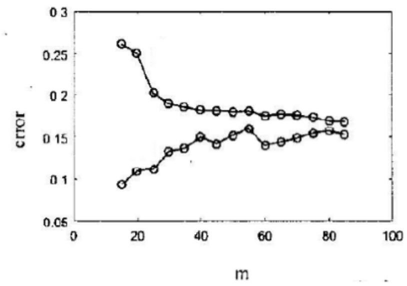
$$m \geq \frac{2 + \log\left(\frac{1}{\delta}\right)}{\epsilon}$$

Section 3: Matching Graphs to Polynomial Kernel Degrees

Dana used SVM with polynomial kernels of degrees $d = 2, 10, 20$. For each degree, she experimented with 15 to 85 training samples, in increments of 5. The graphs below show the train and test errors for each degree, but the graphs are not labeled.

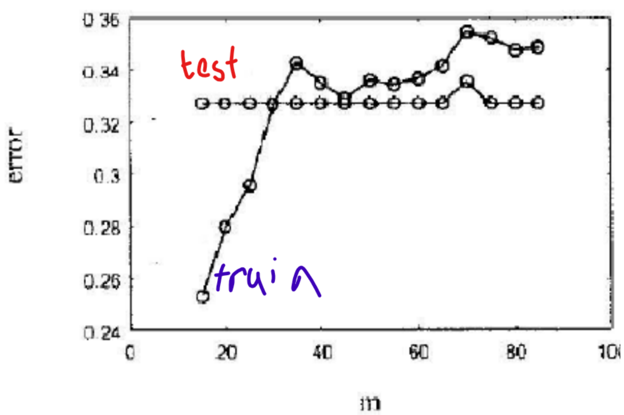


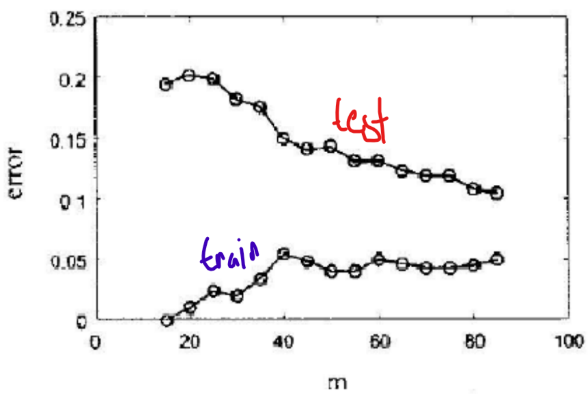


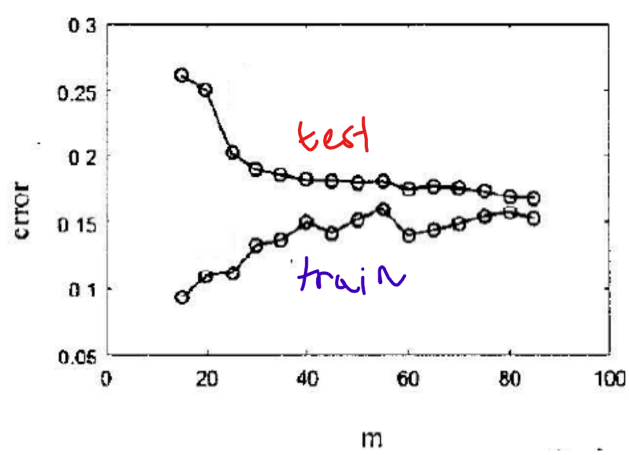


Solution

The graphs are matched as follows:
1. **Left graph**: $d = 2$
2. **Center graph**: $d = 20$
3. **Right graph**: $d = 10$




test
train
error
m
0.25
0.2
0.15
0.1
0.05
0
0
20
40
60
80
100


test
train
error
m
0.3
0.25
0.2
0.15
0.1
0.05
0
20
40
60
80
100