CS 5350/6350: Machine Learning
Homework 3 Solutions

Ahmad Rafi Faqiri

October 30, 2024

1. Paper Problems [36 points + 15 bonus]

1.1 Linear Classifier Margin Calculation

Given a linear classifier for 2-dimensional features, where the classification boundary (hyperplane) is defined as:

$$2x_1 + 3x_2 - 4 = 0$$

We need to determine if the hyperplane has a margin for two datasets provided, and if so, calculate the margin.

(a) Dataset 1 (4 points)

The points in Dataset 1 are:

x_1	x_2	label
1	1	1
1	-1	-1
0	0	-1
-1	3	1

To check if the hyperplane has a margin, we calculate the distance from each point to the hyperplane:

$$d = \frac{|2x_1 + 3x_2 - 4|}{\sqrt{2^2 + 3^2}} = \frac{|2x_1 + 3x_2 - 4|}{\sqrt{13}}$$

Calculating the distance for each point:

- Point (1, 1): $d = \frac{1}{\sqrt{13}}$
- Point (1, -1): $d = \frac{5}{\sqrt{13}}$
- Point (0, 0): $d = \frac{4}{\sqrt{13}}$
- Point (-1, 3): $d = \frac{3}{\sqrt{13}}$

The margin is the minimum of these distances:

$$\text{Margin} = \min\left(\frac{1}{\sqrt{13}}, \frac{5}{\sqrt{13}}, \frac{4}{\sqrt{13}}, \frac{3}{\sqrt{13}}\right) = \frac{1}{\sqrt{13}} \approx 0.277$$

(b) Dataset 2 (4 points)

The points in Dataset 2 are:

x_1	x_2	label
1	1	1
1	-1	-1
0	0	-1
-1	3	1
-1	-1	1

To calculate the margin, we compute the distance d from each point to the hyperplane:

$$d = \frac{|2x_1 + 3x_2 - 4|}{\sqrt{13}}$$

Calculating the distance for each point in Dataset 2:

- Point (1, 1) with label +1: $d = \frac{|2(1)+3(1)-4|}{\sqrt{13}} = \frac{1}{\sqrt{13}}$
- Point (1, -1) with label -1: $d = \frac{|2(1)+3(-1)-4|}{\sqrt{13}} = \frac{5}{\sqrt{13}}$
- Point (0, 0) with label -1: $d = \frac{|-4|}{\sqrt{13}} = \frac{4}{\sqrt{13}}$
- Point (-1, 3) with label +1: $d = \frac{3}{\sqrt{13}}$
- Point (-1, -1) with label +1: $d = \frac{|2(-1)+3(-1)-4|}{\sqrt{13}} = \frac{9}{\sqrt{13}}$

The margin is the minimum of these distances. However, note that point $(-1, -1)$ has a positive label but lies on the side of the hyperplane opposite to its label, indicating that this dataset is not linearly separable by this hyperplane. Thus, a consistent margin does not exist for Dataset 2.

2. Dataset Margin Analysis

(a) Dataset 3 (4 points)

The points in Dataset 3 are:

x_1	x_2	label
-1	0	-1
0	-1	-1
1	0	1
0	1	1

This dataset is linearly separable, as all negative examples can be separated from positive examples by a line. The maximum margin is achieved by choosing the optimal hyperplane that maximizes the distance between the closest points of opposite classes. Calculating distances, we find that a linear classifier can be placed such that it maximizes the separation between the classes.

(b) Dataset 4 (4 points)

The points in Dataset 4 are:

x_1	x_2	label
-1	0	-1
0	-1	1
1	0	-1
0	1	1

This dataset is not linearly separable because it is not possible to draw a single line that separates all positive examples from all negative examples. Hence, a margin does not exist.

3. Bonus: Mistake Bound Theorem for Perceptron (5 points)

Given a vector $u \in R^n$ and a positive constant γ such that for each training example (x_i, y_i):

$$y_i(u^\top x_i) \geq \gamma$$

the upper bound for the number of mistakes made by the Perceptron algorithm is:

$$\frac{R^2}{\gamma^2}$$

where $R = \max_i \|x_i\|$. This bound ensures that the Perceptron converges within a finite number of steps for linearly separable data.

4. Upper Bound on Mistakes for Perceptron Learning a Disjunction [10 points]

Given a disjunction:

$$f(x_1, x_2, \ldots, x_n) = \neg x_1 \vee \cdots \vee \neg x_k \vee x_{k+1} \vee \cdots \vee x_{2k}$$

where $2k < n$, we aim to derive an upper bound for the number of mistakes made by the Perceptron algorithm when learning this function.

For a disjunction, the Perceptron's goal is to find a linear boundary that correctly classifies each input vector in the training set, which consists of all 2^n Boolean vectors.

1. Margin Assumption: Each vector x can be viewed as being correctly classified with some margin γ. Since the Perceptron converges after making at most $\frac{R^2}{\gamma^2}$ mistakes (where R is the maximum norm of any x), we can estimate an upper bound using this property.

2. Norm Calculation: For Boolean inputs, $R = \sqrt{n}$ since each component $x_i \in \{0, 1\}$, making the Euclidean norm $\|x\| \leq \sqrt{n}$.

3. Margin Calculation: Given the construction of $f(x)$ as a disjunction, the smallest margin γ will depend on the closest point to the decision boundary in weight space. For the sake of a worst-case estimate, we assume γ is relatively small and treat it as a constant.

Thus, the upper bound for the number of mistakes is:

$$\frac{R^2}{\gamma^2} = \frac{n}{\gamma^2}$$

5. Linear Classifiers in a Plane and Shattering [10 points]

To prove that linear classifiers in a plane cannot shatter any 4 distinct points, we use the concept of the VC-dimension.

1. VC-Dimension: A set of points is said to be shattered by a hypothesis class if every possible labeling of the points can be realized by some hypothesis in the class. The VC-dimension of linear classifiers in a 2-dimensional plane is 3.

2. **Proof by Contradiction**: - Assume that 4 points can be shattered by a linear classifier. - For 4 points to be shattered, the classifier must correctly separate each of the $2^4 = 16$ possible labelings. - However, in 2D space, if 4 points are arranged such that no three are collinear, they form a convex quadrilateral, which cannot be shattered because there exist labelings that cannot be linearly separated (e.g., labeling two adjacent points as positive and the other two as negative).

Thus, no linear classifier can shatter 4 distinct points in a plane, proving that the VC-dimension is 3.

6. Bonus: Hypothesis Space of Rectangles in a Plane [10 points]

Consider an infinite hypothesis space H consisting of all rectangles in a plane. Each rectangle acts as a classifier, where all points inside the rectangle belong to one class and those outside belong to another.

1. VC-Dimension of Rectangle Classifiers: - To determine the VC-dimension, we find the maximum number of points that can be shattered. - Four points arranged in a general position (such as the vertices of a rectangle) can be shattered by selecting different subsets within or outside the rectangle.

2. Shattering 4 Points: - For four points in general position, any subset of points can be enclosed by a rectangle. This includes selecting subsets like two diagonal points, three corners, etc. - Thus, a rectangle hypothesis can shatter 4 points but not 5, as it is impossible to select arbitrary subsets of five points that can all be separated by a rectangle.

Therefore, the VC-dimension of the hypothesis space of rectangles in a plane is 4.