

Homework 1 Solutions

Math 198: Math for Machine Learning

1 Demographics

1. What year are you in and what is your major?
2. Which courses in the Math department have you taken prior to this course?
3. What courses in the CS or Data Science department have you taken prior to this course?
4. What courses in the Statistics department have you taken prior to this course?
5. What courses in the EE department have you taken prior to this course?
6. Have you taken any other courses which you believe are relevant to this one? If so, which?
7. Are you planning on taking CS 189 or another machine learning course?
8. What led you to enroll in this course, and what are you hoping to get out of it?

2 Perceptrons

1. Suppose we are working with two-dimensional data, and have the following datapoints:
Class A: $\mathbf{x}_1 = [1, 1]$, $\mathbf{x}_2 = [4, 4]$
Class B: $\mathbf{x}_3 = [1, -2]$, $\mathbf{x}_4 = [4, 1]$
 - (a) By observation, determine a decision boundary for this data.
 $x_{i2} = x_{i1} - 1$ is one of many possible decision boundaries. The associated weight vector is $\mathbf{w} = [-1, 1]$ and the bias is $b = 1$.
 - (b) Using the file hw1.py, run the perceptron algorithm with learning rate $r = 0.1$ on this data, and report your final values for \mathbf{w} , b , and the decision boundary determined by the algorithm. (To run the file, run `python hw1.py` from your terminal.)
The final values were $\mathbf{w} = [-\frac{1}{5}, \frac{7}{10}]$ and $b = \frac{1}{10}$; this corresponds to the decision boundary $x_{i2} = \frac{2}{7}x_{i1} - \frac{1}{7}$.
2. Can a perceptron be trained to learn the one-bit XOR operation, using the input values as features? Why or why not? (For those unfamiliar, $a \text{ XOR } b$ is true if and only if $a \neq b$.)
No, as any training data for XOR would not be linearly separable. This issue is illustrated in the figure in Note 1, on the right side – taking the values of the two inputs a and b as features and the output as the class label, we get the exact same graph when plotting the four possible input values to one-bit XOR.
3. Prove that the perceptron algorithm will not converge if the data is not linearly separable (assuming the learning rate is not 0).
Proof by contradiction. Suppose the algorithm did converge on data which is not linearly separable. Then, for some time t , we have $\mathbf{w}^{t+1} = \mathbf{w}^t$ and $b^{t+1} = b^t$. So, for all datapoints x_j , we have

$$w_i^{t+1} = w_i^t + r(\delta_j - y_j^t)x_{ji}$$

and

$$b^{t+1} = b^t + r(\delta_j - y_j^t)$$

This in turn implies that $(\delta_j - y_j^t) = 0$ for all datapoints x_j . But if this is true, then every datapoint is correctly classified by the decision boundary outputted by the perceptron. This, in turn, implies that the data is linearly separable, a contradiction; therefore, the algorithm will not converge on data which is not linearly separable. \square