Statistical Pattern Recognition
Assignment 6

Ali Gholami

Department of Computer Engineering & Information Technology
Amirkabir University of Technology

https://aligholamee.github.io
aligholami7596@gmail.com

Abstract

Keywords.

1 LDF Based on Perceptron Learning Rule

Recall that a perceptron learns a linear classifier with weight vector W. It predicts:

$$\hat{y} = sign(W^T x_t)$$

assuming here that $\hat{y} \in \{+1, -1\}$. When the perceptron makes a mistake, it updates the weights using the formula:

$$W^{t+1} = W^t + y_t x_t$$

Imagine that we have $x_t \in I^2$, and we encounter the following data points:

X[1]	X[2]	y
1	1	1
2	-1	-1
-3	-1	-1
-3	1	1

(a) Starting with $W = [0\ 0]^T$, use the Perceptron Algorithm to learn on the data points in the order from top to bottom. Show the Perceptron's linear decision boundary after observing each data point. Be sure to show which side is classified as positive.

(b) Does our learned perceptron maximize the margin between the training data and the decision boundary? If not, draw the maximum-margin decision boundary.

2 Perceptron in Practice

Start with the truth tables for boolean *AND* and *OR*. Your goal is to write an small program to train a *two-input one-output* Perceptron using this data. Select some random initial weights in the interval $(0, 1)$. Use a learning parameter 0.1. The stopping condition is 50 iterations or *no change in weights from one epoch to the next*, whichever comes first.

(a) Encode *FALSE = 0* and *TRUE = 1* for the input values. Remember that the output of a Perceptron, by definition is $+1$ or -1.

(b) Encode *FALSE = -1* and *TRUE = +1* for the input values. Remember that the output of a Perceptron, by definition is $+1$ or -1.

For each case, please show the evolution in the change of weights by arranging them in a nice tabular format. Remember to include the weight w_0 as a part of the training process and set your x_0 to $+1$ all the time. For each case, please plot the separating hyperplane in the $x_1 - x_2$ plane. Submit your code, result of your run and your comments.

3 Support Vector Machine

Consider a dataset with 2 points in $1D$:

$$X_1 = 0,\ Y_1 = -1 \quad X_2 = \sqrt{2},\ Y_2 = 1$$

Consider mapping each point to $3D$ using the feature vector $\phi(x) = [1,\ \sqrt{2}x,\ x^2]^T$. Here, the max margin classifier has the form:

$$\hat{W},\ \hat{W}_0 = argmin||W||^2 \tag{3.1}$$

so that

$$y_1(W^T\phi_1(x) + W_0) \geq 1$$
$$y_2(W^T\phi_2(x) + W_0) \geq 1$$

(a) Write down a vector that is parallel to the optimal vector $\hat{W}$. Hint: $\hat{W}$ is prependicular to the decision boundary between the two points in the $3D$ feature space.

(b) What is the value of the margin that is achieved by this $\hat{W}$? Hint: The margin is the distance from each support vector to the decision boundary. Think about the geometry of the points in feature space, and the vector between them.

(c) Solve for $\hat{W}$, using the fact that the margin is equal to $\frac{1}{||\hat{W}||}$.

(d) Solve for $\hat{W}_0$ using your value for $\hat{W}$ and (3.1). Hint: The points will be on the decision boundary, so the inequalities will be tight. A *Tight Inequality* is an equality that is as strict as possible.

4 Hierarchical Clustring

Let $X = \{x_i,\ i = 1, 2, 3, 4, 5\}$, with $x_1 = [1,\ 1]^T$, $x_2 = [2,\ 1]^T$, $x_3 = [5,\ 4]^T$, $x_4 = [6,\ 5]^T$ and $x_5 = [6.5,\ 6]^T$. The pattern matrix of X is

$$D(X) = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 5 & 4 \\ 6 & 5 \\ 6.5 & 6 \end{bmatrix}$$

Use the following methods to cluster the pattern matrix and draw the dendogram for hierarchical models:

(a) Single Linkage

(b) Complete Linkage

(c) Average Linkage

(d) Mean Method