AI1001: Introduction to Modern AI
Homework Assignment 3
Due Date: 21 August 2019

1. The object of study in this problem is the set of perceptrons with two inputs, but *without a bias term*. So the perceptron operates as

$$y = \text{sign}(w_1 x_1 + w_2 x_2).$$

Show that the VC-dimension of this restricted class of perceptrons is two. To do this, you must show two things:

- First, you should find a pair of 2-D vectors for which all $2^2 = 4$ labellings can be achieved by a suitable perceptron of the above type.
- Second, you must show that for *every* set of three 2-D vectors, at least one of the $2^3 = 8$ labellings *cannot* be achieved by a perceptron of the above type.

2. This problem involves logistic regression.

- Use logistic regression to find the classifier that maximizes the log-likelihood ratio for the following data set. (The matrix consists of the 2×1 vectors, and the column vector consists of the corresponding labels.)

$$X = \begin{bmatrix} -1 & -0.2 \\ 3.5 & 1.8 \\ 5.3 & 1.2 \\ 2.2 & 1.1 \\ 7.5 & 1.4 \\ 2.3 & -0.3 \\ 4.4 & 1.1 \\ 7.0 & 0.5 \end{bmatrix}, y = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

- Compute the probability of assigning the label 1 to each training vector.
- Plot the vectors $\mathbf{x}_i$ and the optimal separating straight line, and show the probabilities assigned to each training vector.
- Now suppose we have two test inputs, namely $\mathbf{v}_1 = [0.7 \quad 0.2]$ and $\mathbf{v}_1 = [4.7 \quad -0.2]$. Compute the probability with which the logistic regressor assigns the probability of 1 to each test sample.

3. Suppose we have a set of training inputs $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)$, and we train an optimal logistic regression classifier. Now suppose we simply flip the labels on all training inputs, that is, we change y_i to $1 - y_i$ for all i. What happens to the optimal logistic regression classifier? What happens to the probability associated with each training sample?

4. This problem has to do with multi-class classification. Suppose there are three classes. Write down the code matrix for each of the three approaches described in the notes.