Assignment 3

Due: December 1

IMPORTANT: Assignments must be written on a computer, and delivered on moodle.

A. Some probability.
1. [15] Suppose you’re on a game show, and you’re given the choice of three doors: Behind one door is a Ferrari; behind the others, a broken bicycle. You pick a door, say No. 1 [but the door is not opened], and the host, who knows what’s behind the doors, opens another door, say No. 3, which has a broken bicycle. He then says to you, ”Do you want to pick door No. 2?” Is it to your advantage to switch your choice? Prove your answer using conditional probabilities.

B. Perceptron.
1. [10] Design a two-input perceptron that implements the boolean function

   \[ A \text{ OR NOT}(B). \]

Design a two-layer network of perceptrons that output the function

\[ A \text{ XOR } B. \]

2. [5] Consider two perceptrons \( 1 + 2x_1 + x_2 > 0 \) and \( 2x_1 + x_2 > 0 \). Is one more general than the other? (a concept \( C_1 \) is called more general than a concept \( C_2 \) if all data classified as positive from \( C_2 \) are also classified as positive by \( C_1 \).

3. [10] Here is 2 points \((x, y)\) on the 2-dimensional plane, along with their classifications \( c \):
   (i) \((x = 1, y = 1, c = -1)\), (ii) \((x = 3, y = 2, c = 1)\).
Let \( f(x, y) = w_0 + w_1 x + w_2 y \) be a classification function. What is the mean squared error over the input points? Find the parameters \( w_0, w_1, w_2 \) in the classification function \( f(x, y) = w_0 + w_1 x + w_2 y \) so that the mean square error is minimized.

C. Gradient Descent and \( \delta \) rule.
1. [10] Derive a gradient descent training rule for a single unit with output \( o \) where

   \[ o = w_0 + w_1 x_1 + w_1 x_1^2 + \ldots + w_n x_n + w_n x_n^2. \]

D. Hypothesis Evaluation.
1. [10]. Consider a learned hypothesis \( h \) for some boolean concept. When \( h \) is tested on a set of 100 examples, it classifies 83 correctly. What is the standard deviation and the 95 confidence interval for the true error rate \( Error_D(h) \).

2. [15]. What is the difference between inductive learning bias, and estimation bias? Give an example of both cases.
E. Reinforcement Learning.
1. [10] Prove that, that if initialized to 0, the values of $\hat{Q}$ function always increase and always are below the values of the ‘actual’ $Q$ function (chapter 13)

F. VC dimension and PAC Learning.
1. [10] Give a definition for the VC dimension. Why is the VC dimension useful?
2. [25] Consider the space of instances $X$ corresponding to all points in the $x,y$ plane. Give the VC dimension of the following hypotheses spaces:
   1. $H_r$, the set of all rectangles in the plane.
   2. $H_c$, the set of circles in the plane. Points inside the circle are classified as positive.
   3. $H_t$, the set of triangles in the plane. Points inside the triangle are classified as positive.