AI 305 Introduction to Machine Learning

Homework1

Question 1:

Suppose we have a data set with five predictors, X1 = GPA, X2 = IQ, X3 = Level (1 for College and 0 for High School), X4 = product between GPA and IQ, and X5 = product between GPA and Level. The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model, and estimate $\beta_0 = 50$; $\beta 1 = 20$; $\beta 2 = 0.07$; $\beta 3 = 35$; $\beta 4 = 0.01$; $\beta 5 = -10$.

a. Predict the salary of a college graduate with IQ of 110 and a GPA of 4.0.
b. Explain the effect of the low values of $\beta 2$ and $\beta 4$ comparing to the high absolute values of $\beta 1$, $\beta 3$, and $\beta 5$.
c. True or false: (Justify)
    i. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates.
    ii. True or false: For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates provided that the GPA is high enough.

Question 2:

Suppose we collect data for a group of students in a statistics class with variables X1 =hours studied, X2 =undergrad GPA, and Y = receive an A. We fit a logistic regression and produce estimated coefficient,

$\beta_0 = -6$; $\beta_1 = 0.05$; $\beta 2 = 1$.

a. Estimate the probability that a student who studies for 40 h and has an undergrad GPA of 3.5 gets an A in the class.
b. How many hours would the student in part (a) need to study to have a 50% chance of getting an A in the class?

Question 3:

Consider the following set of examples representing dataset for classifying care according set of features into positive or negative cars.

Color	Type	Doors	Tires	Class
Red	SUV	2	Whitewall	+
Blue	Minivan	4	Whitewall	-
Green	Car	4	Whitewall	-
Red	Minivan	4	Blackwall	-
Green	Car	2	Blackwall	+
Green	SUV	4	Blackwall	-
Blue	SUV	2	Blackwall	-
Blue	Car	2	Whitewall	+
Red	SUV	2	Blackwall	-
Blue	Car	4	Blackwall	-
Green	SUV	4	Whitewall	+
Red	Car	2	Blackwall	+
Green	SUV	2	Blackwall	-
Green	Minivan	4	Whitewall	-

Draw the decision tree representing these dataset using:

a. Misclassification rate
b. Entropy and Information gain
c. Gini impurity.

Question 4:

This problem involves hyperplanes in two dimensions.

a. Sketch the hyperplane $1 + 3X1 - X2 = 0$. Indicate the set of points for which $1 + 3X1 - X2 > 0$, as well as the set of points for which $1 + 3X1 - X2 < 0$.
b. On the same plot, sketch the hyperplane $-2 + X1 + 2X2 = 0$. Indicate the set of points for which $-2 + X1 + 2X2 > 0$, as well as the set of points for which $-2 + X1 + 2X2 < 0$.

Question 5:

Maximal margin classifier on a sample data set.

a. We are given n = 7 observations in p = 2 dimensions. For each observation, there is an associated class label. Sketch the observations.

Obs.	X_1	X_2	Y
1	3	4	Red
2	2	2	Red
3	4	4	Red
4	1	4	Red
5	2	1	Blue
6	4	3	Blue
7	4	1	Blue

b. Sketch the optimal separating hyperplane, and provide the equation for this hyperplane (of the form $\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p, =0$).

c. Describe the classification rule for the maximal margin classifier. It should be something along the lines of "Classify to Red if $\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p, >0$, and classify to Blue otherwise." Provide the values for β_0, $\beta 1$, and $\beta 2$.

d. On your sketch, indicate the margin for the maximal margin hyperplane.

e. Indicate the support vectors for the maximal margin classifier.

f. Argue that a slight movement of the seventh observation would not affect the maximal margin hyperplane.

g. Sketch a hyperplane that is not the optimal separating hyperplane, and provide the equation for this hyperplane.

h. Draw an additional observation on the plot so that the two classes are no longer separable by a hyperplane.