COMP3314_2C Machine Learning

Homework1:

1. Consider a Perceptron with 2 inputs and 1 output. Let the weights of the Perceptron be w1=1 and w2=1 and let the bias be w0=−1.5. Calculate the output of the following inputs:(0, 0), (1, 0), (0, 1), (1, 1). (12 points)

Answer (since w0 is the bias unit, the threshold should be 0):

-1, -1, -1, 1

2. Suppose that the following are a set of point in two classes:

- Class1: (1,1), (1,2), (2,1)
- Class2: (0,0), (1,0), (0,1)

(1) Plot them and find the optimal separating line. (10 points)

Answer:

y+x-1.5=0

(2) What are the support vectors, and what is the meaning? (14 points)

Answer:

Support vectors: (1, 0), (0, 1) and (1, 1)

These are the training samples that are closest to this separating hyperplane. They would change the position of the hyperplane if removed.

3. Suppose that the probability of five events are P(first) = 0.5, P(second) = P(third) = P(fourth) = P(fifth) = 0.125. Calculate the entropy and write down in words what this means. (14 points)

Answer:

Entropy=2

Entropy describes the disorder or degree of randomness. An entropy equal to 2 means the event happens randomly.

4. Suppose we collect data for a group of students in a postgraduate machine learning class with features x1 = hours studies, x2 = undergraduate GPA and label y = receive an A. We fit a logistic regression and produce estimated weights as follows: w0=−6, w1=0.05, w2=1.

   (1) Estimate the probability that a student who studies for 40h and has an undergraduate GPA of 3.5 gets an A in the class. (10 points)

   Answer: 0.378

   (2) How many hours would the student in part 1. need to study to have a 50% chance of getting an A in the class? (10 points)

   Answer: 50 hours

5. Given the following dataset:

V	W	X	Y
0	0	0	0
0	1	0	1
1	0	0	1
1	1	0	0
1	1	1	0

Your task is to build a decision tree for classifying variable Y. (You can think of the dataset as replicated many times, i.e. overfitting is not an issue here).

(1) Compute the information gains $IG(Y|V)$, $IG(Y|W)$ and $IG(Y|X)$. Remember, information gain is defined as

$$IG(D_p) = I_G(D_p) - \sum_{j=1}^{m} \frac{N_j}{N_p} I_G(D_j)$$

where

$$I_G(t) = 1 - \sum_{i=1}^{c} p(i|t)^2$$

c is the class number, D_p and D_j are the dataset of the parent and jth child node. I_G is gini impurity. N_p is the total number of samples at the parent node and N_j is the number of samples in the jth child node. (10 points)

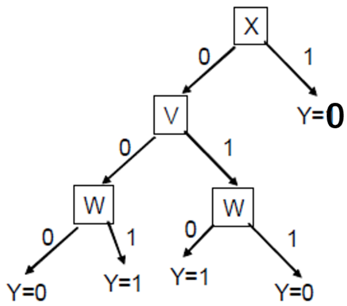
Answer:

$IG(Y|V) = 1/75$, $IG(Y|W) = 1/75$ and $IG(Y|X) = 0.08$

(2) Write down the entire decision tree with gini impurity. (20 points)

Answer:

A full tree is constructed. First we split on X. Given a split on X the information gains for V and W are 0, so we split on either of them (let's say V). Last we split on W (information gain is 1). Figure (a) gives the solution.



(a) Full tree