Artificial Intelligence Spring 2019 Homework 3: Machine Learning

Please follow this format for submission. You may reproduce it in LaTeX or word if you wish.

Name: Vikram Ho

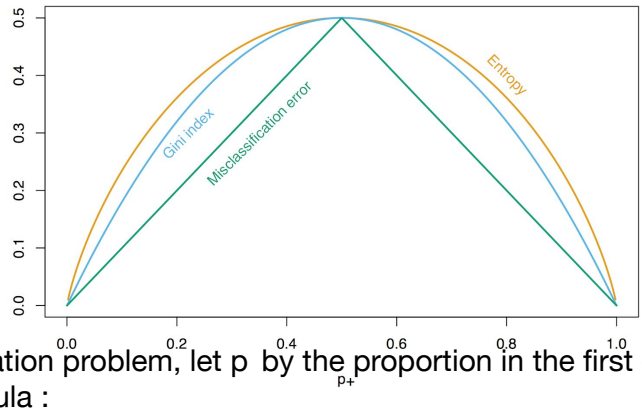
UNI: vwh2107

WRITTEN

Please justify all your answers to receive full credit (unless stated otherwise). There are 3 questions.

Question 1: Decision trees

Explain the reason behind using entropy/gini functions in decision tree rather than the misclassification error? (feel free to research the topic and cite your references). Full credit will be given to clear answers.



For a 2 class classification problem, let p by the proportion in the first class, the each measure has the formula :

Misclassification error : 1 - max(p, 1-p)
Entropy: -p * log - (1-p) * log (1-p)
Gini : 2p * (1-p)

All three measures are similar , but Entrophy and Gini are differentiable, and hence more subject to numerical optimization. For example, from lectures we know that to apply gradient descent which is an optimization technique our functions need to be differentiable, hence in the case of misclassification error, we cannot use a gradient descent

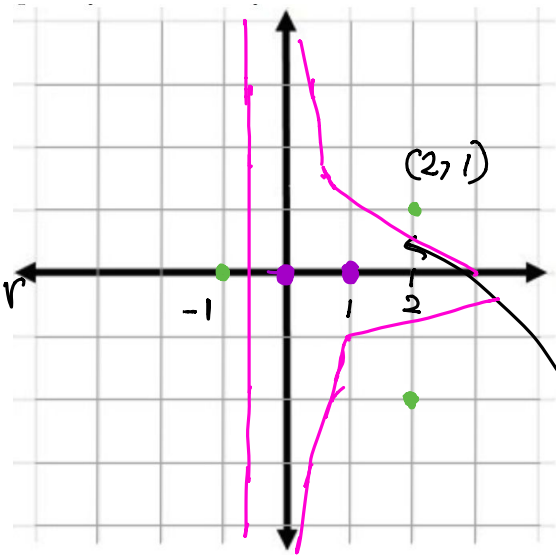
Entrophy and gini are more sensitive to change in the node probabilities than the misclassification error rate.

Question 2: KNN

Suppose we have the following data points:

- Positive: (0, 0) (1, 0)
- Negative: (-1, 0) (2, 1) (2, -2)

1. Draw the graph with the data points along with the decision boundary for **1 Nearest Neighbors**. Briefly explain the shape of your boundary.



• is negative
• is positive

Explanation for decision boundary:

The decision boundary line is calculate using the perpendicular bisectors of a pair of points

$\left(\frac{2+1}{2}, \frac{1+0}{2}\right)$

$\left(\frac{3}{2}, \frac{1}{2}\right)$ (midpoint)

slope of the ⊥ bisector is the negative reciprocal of the slope

$\frac{1-0}{2-1} = \frac{1}{2} \Rightarrow -2$

Equation of bisector

$y = -2x + 3.5$

Do the same for each pair of point and join the perpendicular bisectors',

2. What class does 1NN predict for the new point: (1, 1.01)? Explain why.

The closest point can be (2, 1) or (1, 0)

The distance for new point to (2, 1) is $\sqrt{(2-1)^2 + (1-1.01)^2} \approx 1.00004$

The distance for new point to (1, 0) is $\sqrt{(1-1)^2 + (0-1.01)^2} = 1.01$

Since $1.01 > 1.00004$, i.so the new point is negative.

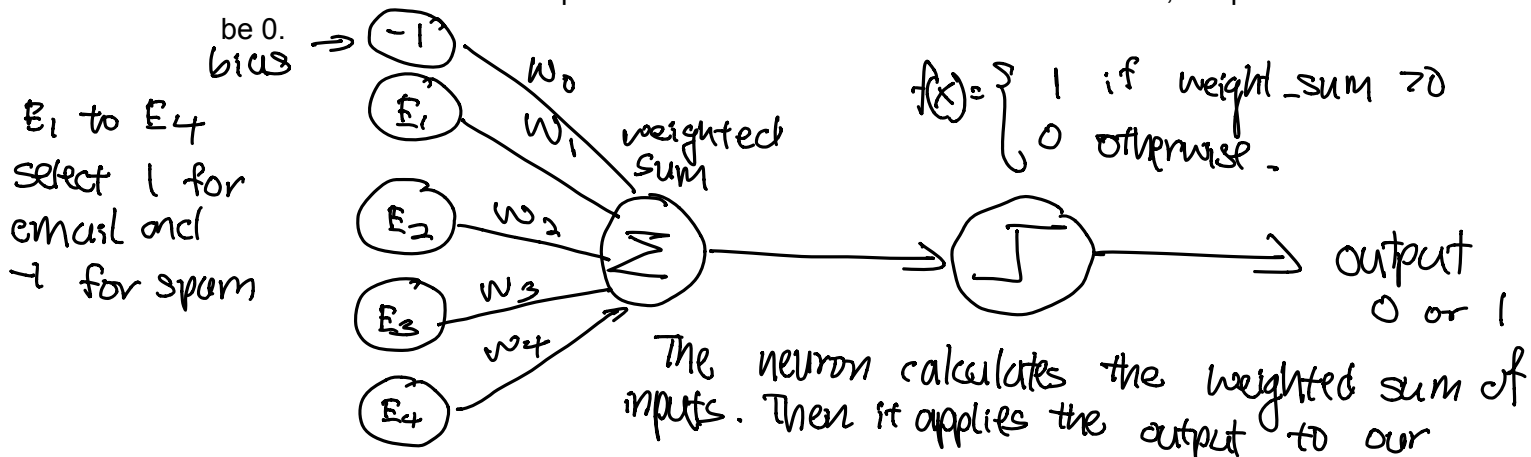
3. What class does 3NN predict for the new point: (1, 1.01)? Explain why.

3NN will predict for the new point as positive. Since it is close to two of the positive points

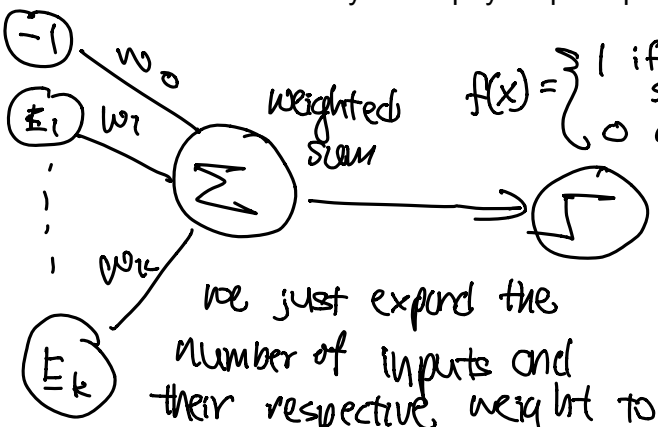
Question 3: Perceptron

We would like to implement a majority voting mechanism for filtering emails. We are given predictions of n different voting entities (e.g., experts, classifiers). The task is to predict whether a text is Spam (class 0) or Email (class 1).

1. Design a perceptron that implements majority voting. For simplicity, we will assume there are 4 experts (named E1, E2, E3, and E4). Draw the network and the weights. Note: a tie happens when two voters vote for spam and two voters vote for email. In this case, the prediction should be 0.



2. How can you adapt your perceptron to a set of n experts? Explain.



Assuming $w_1 = w_2 = w_3 = w_4 = w_5$

If E_1 and E_2 pick Email $\Rightarrow 1$

E_3 and E_4 pick spam $\Rightarrow -1$

$1w_1 + 1w_1 - w_3 - w_4 \cdot 1 = -1 < 0$

so step function predicts 0.

use the same activation step. function

3. Suppose that every individual expert uses completely different information and they all have the same accuracy level (say, 70%) on the given set of examples. Can majority voting improve accuracy in this case on the same set of examples? Explain.

Yes, a major voting system will improve the performance when every individual expert is somewhat independent of each other. For example, if the expert were completely dependent, same classification

errors will occur.
But if the experts were somewhat independent,
ie using different informatics or training set
then a multiple classifer system will
improve the performance