Homework-3

Question 1

You are given the following training examples. Each has only one attribute, and the classification is into positive/negative.

index	x	classification
1	1.0	$+$
2	2.0	$-$
3	4.0	$+$
4	5.0	$+$
5	6.0	$-$
6	7.0	$-$

We would like to evaluate two learning algorithms that use a set S of training examples to classify the example with attribute value of x.

Algorithm A

Let S_p, S_n be the sets of positive and negative examples in S.

If S_p is empty classify x as negative. If S_n is empty classify x as positive.

Otherwise, compute u_p, the mean of the x values in S_p, and u_n, the mean of the x values in S_n.

If x value is closer to u_p than it is to u_n then classify x as positive. Otherwise classify x as negative.

Example: Using all the training examples above we have: $u_p = 3.33$, $u_n = 5$. Therefore, an example with $x = 2.5$ is classified as positive.

Algorithm B

Find the example y with an attribute value nearest to x.

Classify x with the same classification as y.

(If there are two examples nearest to x, one positive and the other negative, classify x as positive.

Example: Using all the training examples above the nearest example to $x = 2.5$ has index 2. Therefore, it is classified as negative.

Part 1.

Use leave-one-out (6-fold cross validation) to estimate the errors of Algorithm A and Algorithm B. Answer:

$$e_A =$$
$$e_B =$$

Part 2.

Use 3-fold cross validation to estimate the errors of Algorithm A and Algorithm B. Run cross-validation three times, on the following three permutations of the data:

1–	$(1, 2, 3, 4, 5, 6)$
2–	$(1, 3, 5, 2, 4, 6)$
3–	$(1, 6, 2, 5, 3, 4)$

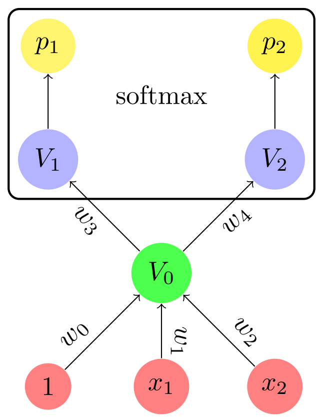
What are the computed errors in each case:

Permutation 1:	$e_A =$	$e_B =$
Permutation 2:	$e_A =$	$e_B =$
Permutation 3:	$e_A =$	$e_B =$

What are the estimates to the error and standard deviation for each algorithm:

Algorithm A:	$e =$	$\sigma =$
Algorithm B:	$e =$	$\sigma =$

Question 2



The above neural network has two inputs. It computes a selection between the two alternatives A, B in terms of two probability outputs. p_1 is the probability that A occurs, and p_2 is the probability that B occurs. The node V_0 is implemented with ReLU. The nodes V_1, V_2 are linear (ADALINE), and they are not connected to a bias. The probabilities p_1, p_2 are computed from the values of V_1, V_2 using softmax.

A.1: Compute the values of all nodes in forward propagation when the network is given the input $\boxed{x_1 = 2,\ x_2 = 7}$, the current weight values are: $\boxed{w_0 = 0,\ w_1 = 0.2,\ w_2 = 0.1,\ w_3 = 0.1,\ w_4 = 1}$, with $\boxed{\text{the desired selection being } \mathbf{A}}$. Use training rate $\boxed{\epsilon = 0.1}$. Your answer should be explicit numeric values for each node.

Question 3

The Adam method uses a recursive method for computing running averages:

$$\overline{X}_0 = 0, \quad \overline{X_t} = \beta \overline{X}_{t-1} + (1-\beta) X_t, \quad \hat{X}_t = \frac{\overline{X}_t}{1-\beta^t}$$

1. Show that if $\beta = 0$ then $\hat{X}_t = X_t$ for all t.

2. Show that if $\beta \to 1$ then $\hat{X}_t \to \frac{1}{t}\sum_{i=1}^{t} X_i$ for all t.

Hint: Use the explicit formula for $\overline{X_t}$, form the limit, and solve it using, for example, L'Hopital's rule.

Question 4

In the table below cases 3,4 are distributions, and cases 1, 2 can be converted into distributions.

case	A	B	C	D
1	1	-2	3	-4
2	1	2	-3	0
3	1	0	0	0
4	1/4	1/4	1/4	1/4

1. Use cross entropy to determine which distribution among 1,2,3 is most similar to 4. **Show your computations.**

Answer: 1 / 2 / 3

2. Use cross entropy to determine which distribution among 1,2,4 is most similar to 3. **Show your computations.**

Answer: 1 / 2 / 4

Question 5

In this question, if you need to compute logarithms use natural basis logarithm (ln).

Consider a deep neural net applied to decide between the following three categories: A, B, C.

1.

What is the one-hot encoding of the category A?

Answer:

The one-hot encoding of the category A is: ____________

2.

Assuming that the true category is A, what is the cross entropy loss of the following estimate, given as the vector $z_1 = (1.0, 2.0, 3.0)$. You must show your computations.

Answer:

The cross entropy loss is ____________

3.

As in Part 2 assume that the true category is A. Consider an algorithm that updates the prediction by adding the value of ϵ to to each prediction value. (ϵ can also be negative.) The new prediction vector is $z_2 = (1 + \epsilon, 2 + \epsilon, 3 + \epsilon)$. Compute a value of ϵ that gives the best prediction according to cross entropy. What is the cross entropy loss for that value of ϵ? You must show your computations.

Answer:

$\epsilon =$

The cross entropy loss is ____________

4.

As in Part 2 assume that the true category is A. Consider an algorithm that updates the prediction by multiplying each value by ϵ. (ϵ can also be negative.) The new prediction vector is $z_3 = (1 \cdot \epsilon, 2 \cdot \epsilon, 3 \cdot \epsilon)$. Compute a value of ϵ that gives the best prediction according to cross entropy. What is the cross entropy loss for that value of ϵ? You must show your computations.

Answer:

$\epsilon =$

The cross entropy loss is ____________