Congratulations! You passed!

Next Item

1 / 1 points

1. If the output of a model is given by $y = f(\mathbf{x}; W)$, then which of the following choices for f are most appropriate when the task is binary classification?

- [ ] Linear

Un-selected is correct

- [x] Logistic sigmoid

Correct
There can be more than one reasonable choice.

- [x] Binary threshold

Correct
There can be more than one reasonable choice.

- [ ] Linear threshold

Un-selected is correct

1 / 1 points

2. After learning using the Perceptron algorithm, how easy is it to express the learned weight vector in terms of the input vectors and the initial weight vector? Assume the input vectors have real-valued components.

- [x] It requires only one integer per training case.

Correct
For every training case, the update to the weight matrix is determined by the output of the perceptron unit, this is 1 bit of information. However, we can also represent the model with an integer that stores whether we added / subtracted or left the weight matrix unchanged when we looked at that example (-1, 0, +1).

- [ ] It requires real numbers.

Un-selected is correct

- [ ] It is impossible.

Un-selected is correct

- [x] It requires one bit per training case.

Correct
For every training case, the update to the weight matrix is determined by the output of the perceptron unit, this is 1 bit of information. However, we can also represent the model with an integer that stores whether we added / subtracted or left the weight matrix unchanged when we looked at that example (-1, 0, +1).

1 / 1 points

3. Suppose we are given three data points:

$$\begin{array}{c} \mathbf{x} \rightarrow t \\ 1,0 \rightarrow 1 \\ 1,1 \rightarrow 1 \\ 0,1 \rightarrow 0 \end{array}$$

Furthermore, we are given the following weight vector (where the bias is set to 0):

$\mathbf{w} = (0, -3)$

Let $||\mathbf{w}^{(t)} - \mathbf{w}^{(t-1)}||_2$ be the distance between the weight vectors at iteration t and iteration $t-1$ of the perceptron learning algorithm. Here, for a given 2D vector $\mathbf{v}$, $||\mathbf{v}||_2 = \sqrt{v_1^2 + v_2^2}$ (this is also called the Euclidean norm). What is the maximum amount by which the weight vectors can change between successive iterations? Note that in this example we are *not* learning the bias.

- ( ) $2\sqrt{2}$
- ( ) 1
- (x) $\sqrt{2}$

Correct
Let's say that at time t we observe that we have misclassified some point $\hat{\mathbf{x}}$ with target $\hat{t}$. Then the learning algorithm will proceed as:

$$\mathbf{w}^{(t)} = \begin{cases} \mathbf{w}^{(t-1)} + \hat{\mathbf{x}} \text{ if } \hat{t} = 1 \\ \mathbf{w}^{(t-1)} - \hat{\mathbf{x}} \text{ if } \hat{t} = 0 \end{cases}$$

In either case, the distance between $\mathbf{w}^{(t)}$ and $\mathbf{w}^{(t-1)}$ will be $||\mathbf{w}^{(t-1)} - \mathbf{w}^{(t-1)} \pm \hat{\mathbf{x}}||_2 = || \pm \hat{\mathbf{x}}||_2 = ||\hat{\mathbf{x}}||_2 \leq \sqrt{2}$ since this is the length of the largest input vector (in this case, $(1, 1)$).

- ( ) 2

1 / 1 points

4. Suppose that we have a perceptron with weight vector $\mathbf{w}$ and we create a new set of weights $\mathbf{w}^* = c\mathbf{w}$ by scaling $\mathbf{w}$ by some positive constant c.

Assume that the bias is zero.

True or false: if the perceptron now uses $\mathbf{w}^*$ instead then it's classification decisions might change (that is, we have moved the classification boundary).

- (x) False

Correct
If the bias term is zero, all of the hyperplanes that represent individual cases go through the origin of weight space. So changing the length of the weight vector without changing its direction cannot change which side of the plane it lies on.

- ( ) True

1 / 1 points

5. Suppose that we have a perceptron with weight vector $\mathbf{w}$ and we create a new set of weights $\mathbf{w}^* = \mathbf{w} + \mathbf{c}$ by adding some constant vector $\mathbf{c}$ to $\mathbf{w}$. Assume that the bias is zero.

True or false: if the perceptron now uses $\mathbf{w}^*$ instead then it's classification decisions might change (that is, we have moved the classification boundary).

- (x) True

Correct
Adding a constant vector can change the direction of the weight vector. This might change the side on which some data points lie.

- ( ) False

1 / 1 points

6. Suppose we are given four training cases:

$$\begin{array}{c} \mathbf{x} \rightarrow t \\ 1,1 \rightarrow 1 \\ 1,0 \rightarrow 0 \\ 0,1 \rightarrow 0 \\ 0,0 \rightarrow 1 \end{array}$$

It is impossible for a binary threshold unit to produce the desired target outputs for all four cases. Now suppose that we add an extra input dimension so that each of the four input vectors consists of three numbers instead of two.

Which of the following ways of setting the value of the extra input will create a set of four input vectors that is linearly separable (i.e. that can be given the right target values by a binary threshold unit with appropriate weights and bias).

- [ ] Make the third value of each input vector be the same as the first value.

Un-selected is correct

- [x] Make the third value of each input vector be the same as the target value for that input vector.

Correct

- [x] Make the third value be 1 for one of the four input vectors and 0 for the other three.

Correct
Imagine the four input vectors as lying at the corners of a horizontal square. If we raise one corner of the square in the third dimension we can now insert a tilted plane that goes beneath two diagonally opposite corners and above the other two diagonally opposite corners.

- [ ] Make the third value of each input vector be the opposite of the first value (i.e. use 1 if the first value is 0 and 0 if the first value is 1)

Un-selected is correct

1 / 1 points

7. Brian wants to use a neural network to predict the price of a stock tomorrow given today's price and the price over the last 10 days. The inputs to this network are price over the last 10 days and the output is tomorrow's price. The hidden units in this network receive information from the layer below, transmit information to the layer above and do not send information within the same layer. Is this an example of a feed-forward network or a recurrent network?

- ( ) Recurrent
- (x) Feed-forward

Correct
Even though Brian's network is modelling a sequence, it is doing this in an entirely feed-forward fashion. Another name for this kind of model is a nonlinear autoregressive process. Recurrent networks are much more powerful for this task and can do a much better job, however they are also more difficult to train.

1 / 1 points

8. Brian and Andy are having an argument about the perceptron algorithm. They have a dataset that the perceptron cannot seem to classify (that is, it fails to converge to a solution). Andy reasons that if he could collect more examples, that might solve the problem by making the data set linearly separable and then the perceptron algorithm will converge. Brian claims that collecting more examples will not help. Which one of them is correct?

- ( ) Andy
- (x) Brian

Correct
If any set A of points is not linearly separable from set B, then adding more examples to either set cannot make them linearly separable.