Question 1

1 / 1 pts

We have examples in the form of 20 boolean variables, $< x_1, x_2, \ldots, x_{20} >$, and know the true function $f(X)$ is in the class of monotone conjunctions. Say we have a "teacher" who knows the true function and must teach the true function through a set of examples; the true function is $y(X) = x_1 \wedge x_2 \wedge x_4 \wedge x_9 \wedge x_{13} \wedge x_{18}$. What is the minimum number of examples that are required to learn this function?

- 20
- Correct! 7
- 6
- 1

Question 2

1 / 1 pts

Consider a function f that we are trying to learn over the feature space $\{x_1, x_2\}$. We are given the following examples:

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

The Perceptron algorithm can correctly identify a hyperplane for f that separates positive from negative examples.

True

Correct!

False

Question 3

1 / 1 pts

Suppose we have a weight vector $w \in R^2$ with input vectors $x_i \in R^2$ and $y_i \in \{-1, 1\}$, let us initialize our 2-dimensional weight vector to be $w = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Also, suppose we only have 2 examples in our dataset: $\left(x_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, y_1 = -1\right)$, $\left(x_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, y_1 = 1\right)$. After training a model based on the Perceptron algorithm on the above dataset over 1 epoch, which option represents the correct final state of the weight vector if the linear threshold function is $\hat{y} = sgn\{w^T \cdot x \geq 0\}$?

Correct!

[-1,1]

[0,-1]

[-1,-1]

[0,-2]

Question 4

1 / 1 pts

We have previously introduced the SGD-LMS algorithm with an update rule of

$w_{t+1} = w_t + c \cdot (target_i - output_i)\, x_i$

The latter part, after the learning rate c, is also called the gradient g_t where t represents the t^{th} update, so the update rule can also be written as

$w_{t+1} = w_t + c \cdot g_t$

Note that we were using a constant learning rate c. Instead, we now change the algorithm and use (1) a per-feature learning rate, and (2) an adaptive learning rate over time. Specifically, the learning rate at the t^{th} update now becomes a vector r_t, of the same dimensionality as w_t. We can then write the weight update rule as:

$w_{t+1} = w_t + r_t^T g_t$

where $r_t\,[j] = \frac{1}{\sqrt{\sum_{k=1}^{t} g_k[j]^2}}$, $j \in [0, \textbf{dimensionality}(r_t))$

Which of the following statement is true?

(The notation $x\,[i]$ represents the i^{th} element in vector x)

Correct!

- (selected) The learning rate for features that are more likely to be activated will be smaller over time
- The learning rate for features that are more likely to be activated will be larger over time

Question 5

1 / 1 pts

Consider a perceptron algorithm performed over three training instances e_1, e_2, e_3 in this order, where e_2 and e_3 are identical.

If the algorithm makes a mistake on e_2 and an update is performed to the weight vector, will the model correctly predict e_3 next?

Yes

No

Correct!

Unknown

Quiz Score: **5** out of 5