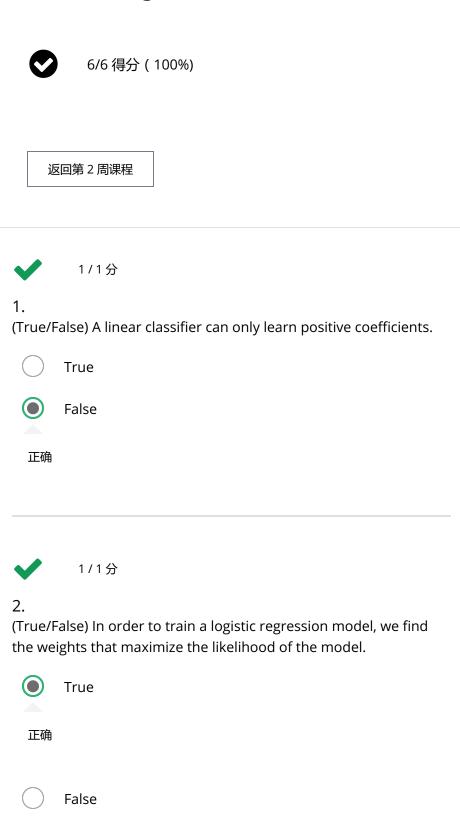
## Learning Linear Classifiers



3.

(True/False) The data likelihood is the product of the probability of the inputs  ${\bf x}$  given the weights  ${\bf w}$  and response y .



正确



1/1分

4.

Questions 4 and 5 refer to the following scenario.

Consider the setting where our inputs are 1-dimensional. We have data

x	y
2.5	+1
0.3	-1
2.8	+1
0.5	+1

and the current estimates of the weights are  $w_0=0$  and  $w_1=1$  . ( $w_0$  : the intercept,  $w_1$  : the weight for x ).

Calculate the likelihood of this data. Round your answer to 2 decimal places.

0.23

正确回答

$$P(y_1 = +1|x_1, w)P(y_2 = -1|x_2, w)P(y_3 = +1|x_3, w)P(y_4 = +1|x_4, w)$$

$$= \frac{1}{1 + e^{-2.5}} \frac{e^{-0.3}}{1 + e^{-0.3}} \frac{1}{1 + e^{-2.8}} \frac{1}{1 + e^{-0.5}}$$

$$= 0.230765 \cdots$$



1/1分

5

Refer to the scenario given in Question 4 to answer the following:

Calculate the derivative of the log likelihood with respect to  $w_1$  . Round your answer to 2 decimal places.

正确回答

$$\frac{\partial \ell(\mathbf{w})}{\partial \mathbf{w}_1} = \sum_{i=1}^4 h_1(\mathbf{x}_i) \left( \mathbf{1}[y_i = +1] - P(y_i = +1|\mathbf{x}_i, \mathbf{w}) \right) 
= 2.5 \left( 1 - \frac{1}{1 + e^{-2.5}} \right) + 0.3 \left( 0 - \frac{1}{1 + e^{-0.3}} \right) 
+ 2.8 \left( 1 - \frac{1}{1 + e^{-2.8}} \right) + 0.5 \left( 1 - \frac{1}{1 + e^{-0.5}} \right) 
= 0.366591 \cdots$$



1/1分

6.

Which of the following is true about gradient ascent? Select all that apply.



It is an iterative algorithm



	It only updates a few of the parameters, not all of them
未选	择的是正确的
<u> </u>	It finds the maximum by "hill climbing"
正确	

