Quiz Submissions - Quiz 6

Mike Gao (username: zenghao.gao@mail.mcgill.ca)

Attempt 1

Written: Oct 26, 2020 1:30 PM - Oct 26, 2020 3:13 PM

Submission View

Released: Oct 25, 2020 8:56 PM

Logistic Regression

Question 1 — **1 / 1 point**

Select **all** correct statements:

- ✓ [ ] The gradient descent solution to the logistic regression model might converge to a local optimum, and fail to find a global optimal solution.
- ✓ [ ] In logistic regression, we model the odds ratio p/(1-p) as a linear function
- ✓ [ ] Logistic regression is a regression method to estimate class posterior probabilities
- ✓ [x] The cost function of logistic regression is convex.

Question 2 — **1 / 1 point**

For the collected data of a group of students in a class with variables X1 = hours studied, X2= GPA and Y = receives an A. We fit a logistic regression with the coefficients w_0 = -6, w_1= 0.05, w_2 = 1.

Estimate the probability that a student who studies for 40 hours and has a GPA of 3.5 **does not get an A** in the class? *Answer by changing probability to percentage (%) and round off the answer to the nearest integer.*

Answer:

62 ✓

Hide Feedback

$$\begin{aligned}\hat{y} &= \frac{\exp^{w_0+w_1X_1+w_2X_2}}{1+\exp^{w_0+w_1X_1+w_2X_2}}\\ &= \frac{\exp^{-6+0.05\times 40+1\times 3.5}}{1+\exp^{-6+0.05\times 40+1\times 3.5}}\\ &= \frac{\exp^{-0.5}}{1+\exp^{-0.5}}\\ &= 37.75\%\end{aligned}$$

$$P(\text{Does not receives A}) = 1 - P(\text{Receives A}) = 100\% - 37.75\% =\sim 62\%$$

Question 3 — 1 / 1 point

For the collected data of a group of students in a class with variables X1 = hours studied, X2= GPA and Y = receives an A. We fit a logistic regression with the coefficients w_0 = -6, w_1= 0.05, w_2 = 1.

Same as the previous question, how many hours would the student need to study with 3.5 GPA **to have a 50% chance of getting an A** in the class?

Answer:

50 ✔

Hide Feedback

$$\begin{aligned}\hat{y} &= \frac{\exp^{w_0+w_1X_1+w_2X_2}}{1+\exp^{w_0+w_1X_1+w_2X_2}}\\ \frac{1}{2} &= \frac{\exp^{-6+0.05x_1+3.5}}{1+\exp^{-6+0.05x_1+3.5}}\\ \exp^{-6+0.05x_1+3.5} &= 1\\ \exp^{-2.5+0.05x_1} &= 1\\ -2.5+0.05x_1 &= 0\\ x_1 &= 50\end{aligned}$$

Question 4 — 1 / 1 point

The logit function is also called log-odds. Consider the classification problem of the Default dataset, where we predict whether an individual will default (Y the target) on his or her credit card payment based on the monthly credit card balance (X the input).

What fraction of people with an odds ratio of 0.37 of defaulting on their credit card payment will in fact default? *Answer by changing probability to percentage (%) and round off the answer to the nearest integer.*

Answer:

27 ✓

Hide Feedback

$$\hat{y} = P(default = yes|balance)$$

$$logit(p) = log(\frac{p}{1-p})$$

$$odds = \frac{p}{1-p}$$

$$\frac{\hat{y}}{1-\hat{y}} = 0.37$$

$$\hat{y} = 27\%$$

Linear Regression

Question 5 — 1 / 1 point

When you find your linear regression model is not converging, you should increase the learning rate such that the weight is updated faster and thus the model will converge faster.

- ○ True
- ✓ ◉ False

Question 6 — 1 / 1 point

Let X and Y be real-valued random variables and Y is generated from X as follows:

$$\epsilon \sim N(0, \sigma^2)$$

$$Y = aX + \epsilon$$

Here, ϵ is the independent variable which is drawn from Gaussian Distribution with 0 mean and σ standard deviation. This is a single feature linear regression model, where “a” is the only weight parameter. The conditional probability of Y has distribution p(Y |X, a) ~ N(aX, σ^2), so it can be written as:

$$p(Y|X,a) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2}(Y - aX)^2\right)$$

Consider, we have a dataset with "n" training example where i = 1...n. Select ALL which correctly represents the maximum likelihood estimation of parameter "a".

- [ ] $$\arg\max_a \sum_i \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{1}{2\sigma^2}(Y_i - aX_i)^2)$$
- [x] $$\arg\max_a \prod_i \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{1}{2\sigma^2}(Y_i - aX_i)^2)$$
- [ ] $$\arg\max_a \frac{1}{2} \sum_i (Y_i - aX_i)^2$$
- [x] $$\arg\min_a \frac{1}{2} \sum_i (Y_i - aX_i)^2$$

Gradient Descent

Question 7 — 1 / 1 point

For linear regression, if we minimize the sum of squared error using the gradient descent, we can get multiple local minima solutions.

- ( ) True
- (•) False ✓

Question 8 — 1 / 1 point

Consider a classification problem with two features x_1 and x_2. In this question, we will walk through a single step of **stochastic gradient descent on logistic regression with cross entropy loss**. Let's say for a given example, target value Y=1 is given and the two features values are x_1 =3 and x_2=2.

Let's assume that initial weights and bias (we call then together as a vector w with three dimension w1, w2 and b) is set to 0 and the learning rate is alpha = 0.1:

w1 = w2= b = 0

alpha = 0.1 (learning rate)

What would the vector w=[w1, w2, b] be after one step of update of w?

- ◯ [w1=0, w2=0, b=0]
- ✔ ◉ [w1=0.15, w2=0.1, b=0.05]
- ◯ [w1=-1.5, w2=-1.0, b=-0.5]
- ◯ [w1=-0.15, w2=-0.1, b=-0.05]

Hide Feedback

$$L_{CE}(w,b) = -[y(\log\sigma(w.x+b)) + (1-y)\big(\log(1-\sigma(w.x+b))\big)]$$

$$\nabla_w = \begin{bmatrix} \frac{\partial L_{CE}(w,b)}{\partial w_1} \\ \frac{\partial L_{CE}(w,b)}{\partial w_2} \\ \frac{\partial L_{CE}(w,b)}{\partial b} \end{bmatrix} = \begin{bmatrix} (\sigma(w.x+b)-y)x_1 \\ (\sigma(w.x+b)-y)x_2 \\ (\sigma(w.x+b)-y) \end{bmatrix} = \begin{bmatrix} (\sigma(0)-1)x_1 \\ (\sigma(0)-1)x_2 \\ (\sigma(0)-1) \end{bmatrix} = \begin{bmatrix} -1.5 \\ -1.0 \\ -0.5 \end{bmatrix}$$

$$w^1 = w^0 - \alpha \begin{bmatrix} -1.5 \\ -1.0 \\ -0.5 \end{bmatrix}$$

$$w^1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - 0.1 \begin{bmatrix} -1.5 \\ -1.0 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 0.15 \\ 0.1 \\ 0.05 \end{bmatrix}$$

Question 9 **1 / 1 point**

For the linear regression problem with cost function as:

$$J(w) = \frac{1}{2}\sum_{i=1}^{n}(y_{(i)} - wx_{(i)})^2$$

where i =(1 ..n) are the training examples with “y” are the real-valued target, “x” is the real-valued feature and “w” as the only weight parameter. **Find the optimal value of w which minimizes the above cost** for given training data below with 3 examples. *Answer up to two decimal points.*

i	x (feature)	y (target)
1	2	2
2	4	1
3	6	1

Answer:

0.25 ✓

Hide Feedback

$$\nabla_w J(w) = \sum_{i=1}^{n}(y_{(i)} - wx_{(i)})(-x_{(i)})$$

$$0 = \sum_{i=1}^{n}(wx_{(i)} - y_{(i)})(x_{(i)})$$

$$0 = \sum_{i=1}^{n}(wx_{(i)}^2 - y_{(i)}x_{(i)})$$

$$w^* = \frac{\sum_{i=1}^{n} y_{(i)}x_{(i)}}{\sum_{i=1}^{n} x_{(i)}^2}$$

$$w^* = \frac{(2\times 2) + (4\times 1) + (6\times 1)}{2^2 + 4^2 + 6^2}$$

$$w^* = 0.25$$

Attempt Score: 9 / 9 - 100 %

Overall Grade (highest attempt)**:** 9 / 9 - 100 %

Done