

Discussion Session Problems 2

01/22/2026

1. Which of the following methods do we use to best fit the data in Logistic Regression?
 - A) Least Squares Error
 - B) Maximum Likelihood
 - C) Both A and B
2. You are given a coin whose probability of landing on Heads is p . We toss the coin 10 times and observe 7 Heads. What is the most likely value of p ?
 - A) 7/10
 - B) 5/10
 - C) 3/10
3. Which of the following evaluation metrics does **not** make sense if applied to logistic regression output to compare with the target?
 - A) Accuracy
 - B) Log loss
 - C) Mean Squared Error
4. What can you say about feature normalization in logistic regression?
 - A) It is good practice but not strictly required
 - B) It is required
 - C) It is bad practice and should not be performed
 - D) None of the above
5. Consider m independent and identically distributed (i.i.d.) random variables x_1, \dots, x_m drawn from a Geometric distribution with parameter p . In other words, $x_i \sim Geo(p)$ for all $1 \leq i \leq m$.
 - (a) Derive the maximum likelihood estimate (MLE) for p .
 - (b) Given $x = \{4, 3, 4, 2, 7\}$, compute \hat{p}_{MLE} . Namely, the value of p in the Geometric distribution that would maximize the likelihood of these observations.
6. In this problem, we simultaneously estimate the difficulty of problem set questions and the skill level of each student.

Consider a set of 200 students and 10 questions, where each student answers each question. Let S_{ij} be an indicator variable such that

$$S_{ij} = \begin{cases} 1 & \text{if student } i \text{ answers question } j \text{ correctly} \\ 0 & \text{otherwise} \end{cases}$$

We assume that the probability that student i answers question j correctly is

$$p_{ij} = \sigma(a_i - d_j),$$

where:

- $\sigma(\cdot)$ is the sigmoid function,
- a_i represents the ability of student i ,
- d_j represents the difficulty of question j .

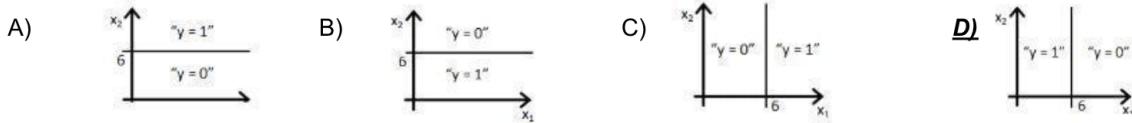
We use Maximum Likelihood Estimation (MLE) to estimate all parameters.

- Write the log-likelihood for a single response S_{ij} in terms of p_{ij} . Hint: logistic regression also assumes that its output is a probability of a binary event.
- Compute the partial derivative of the log-likelihood for a single response S_{ij} with respect to a_i .
- Compute the partial derivative of the log-likelihood for a single response S_{ij} with respect to d_j .
- Explain briefly how the parameters can be estimated using derivatives of log-likelihood with respect to those parameters.

7. Suppose a logistic regression model is

$$f_{w,b}(x) = g(w_1x_1 + w_2x_2 + b)$$

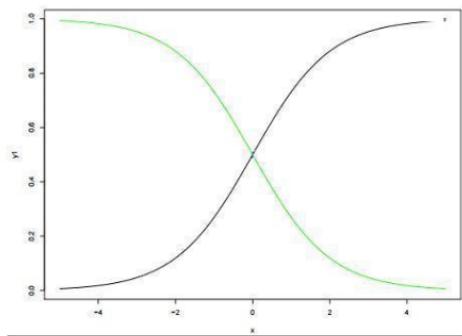
with $b = 6$, $w_1 = 0$, $w_2 = -1$, and g is the sigmoid function. Which figure corresponds to the decision boundary?



8. If w is constant, what is the slope of the logistic function at $x = 0$ for $g(wx)$?

- w
- $w/4$
- $1/4$
- w^2

- Consider $f_{w,b} = g(w_1X + b)$, where w_1 is the coefficient and b is the intercept. Below are two different logistic models with different values for w_1 and b . Which of the following statement(s) is true? The model represented by the green line starts on top.



- w_1 for Green is greater than Black
- w_1 for Green is smaller than Black
- w_1 for both models is the same
- We cannot say for certain the relationship between the w_1 for Green and the w_1 for Black.

10. What can you say about regularized vs. non-regularized logistic regression?
 - It will perform better on the training set
 - We expect it to perform better on the training set
 - It will perform better on the test set
 - We expect it to perform better on the test set
11. In logistic regression, what happens to the predicted probability $p = \sigma(z)$ as $z \rightarrow +\infty$ and $z \rightarrow -\infty$?
 - $p \rightarrow 0$ as $z \rightarrow +\infty$, $p \rightarrow 1$ as $z \rightarrow -\infty$
 - $p \rightarrow 1$ as $z \rightarrow +\infty$, $p \rightarrow 0$ as $z \rightarrow -\infty$
 - $p \rightarrow 0.5$ in both cases
 - p oscillates between 0 and 1
12. Suppose a dataset is perfectly linearly separable. What can happen to the magnitude of the weights learned by unregularized logistic regression?
 - The weights converge to zero
 - The weights converge to a finite value
 - The weights can grow arbitrarily large
 - The algorithm fails immediately
13. Suppose a logistic regression model outputs probabilities very close to 0 or 1 for most training examples. Which of the following is the most likely explanation?
 - The model is underfitting
 - The model weights have very small magnitude
 - The model is highly confident in its predictions
 - The learning rate is too small
14. Which of the following best explains why we take the logarithm of the likelihood in logistic regression?
 - To make the likelihood larger
 - To convert products over data points into sums
 - To ensure predictions lie between 0 and 1
 - To remove the need for regularization
15. Consider a binary classification problem with a very imbalanced dataset (e.g., 99% of labels are 0). Which issue is most likely to arise when training logistic regression?
 - The model cannot be trained using gradient descent
 - Accuracy may be misleading as an evaluation metric
 - The sigmoid function becomes non-differentiable
 - Regularization has no effect

16. Which of the following statements about the intercept (bias) term in logistic regression is true?

- A) It controls the slope of the decision boundary
- B) It shifts the decision boundary without changing its orientation
- C) It is unnecessary if features are normalized
- D) It must always be regularized

17. Suppose we increase the strength of L_2 regularization in logistic regression. Which of the following effects do we expect?

- A) Training loss decreases
- B) Weight magnitudes decrease
- C) The decision boundary becomes more complex
- D) The model fits noise more closely