Week 04 Quiz: Logistic Regression

Name SID

November 27, 2020

Exercise 1

In this exercise, you will compute the predictions and the cross entropy loss of a logistic regression model by hand, so that you will get familiar with the definitions and properties of logistic regression.

Suppose we are given the data set $\{((x_i, z_i), y_i)\}$ for $i = 1, ..., 5$. There are 5 data points in total, each data point has a feature vector $(x, z) \in \mathbb{R}^2$ and label value $y \in \{0, 1\}$.

x_i	-5	-4	-3	-2	-1
z_i	1	2	3	4	5
y_i	0	0	1	0	1

Suppose our logistic regression model for binary classification is given by

$$f_\theta(x, z) = \frac{1}{1 + \exp\{-(\theta_1 x + \theta_2 z)\}}$$

, where $\theta_1 = -1, \theta_2 = 2$. You should remember from the notes that there's no closed form formula to optimize the model parameter θ. Now we want to evaluate performance of this model on the given data set.

1. Compute predictions of this logistic regression model on each data point in this given data set. You may leave your answer as unsimplified fractions
2. Compute the cross entropy loss of this logistic regression model on the whole dataset. You may refer to your computed numerical answers in previous part.

Exercise 2

In this question we want to work with different evaluation metrics of binary classifiers. Suppose we have a binary classifier that classifies the given data set with 10 points in total as follows.

y_i	1	0	1	0	1	0	1	0	1	1
$\hat{y}_i$	0	0	1	0	0	1	1	0	1	0

1. Draw confusion matrix of this logistic regression model on the given dataset
2. Find the accuracy, precision, recall, true positive rate, and false negative rate of this logistic regression model
3. Comment on the performance of this classifier. Is it better or worse than a random guessing classifier? Justify your answer, or leave it blank if there's not enough information.

Exercise 3

In this question, we'll compare the different classification models you have seen so far on linearly separable data.

1. First we consider perceptrons. Recall that perceptrons works as intended only on linearly separable data for binary classification problems.

   Describe the decision boundary found by perceptron algorithms on linearly separable data in binary classification. What are its properties and what's special?

2. Second we discuss the hard-margin SVM on linearly separable data. Describe the decision boundary found by the hard-margin SVM on linearly separable data in binary classification. Compare its shape and properties to perceptrons. What are some advantages and disadvantages of hard-margin SVM?

3. Third consider logistic regression on linearly separable data. You should remember that the optimal weight vector of logistic regression minimizes the cross entropy loss on the training dataset.

   (a) For linearly separable data, what is the value of cross entropy loss? What can you say about the decision boundary found by a logistic regression model?

   (b) Logistic regression given by

   $$f_\theta(x) = \frac{1}{1 + \exp(-\theta^T x)}$$

   is completely determined by weight vector θ. If you want to force cross entropy loss as close to the value you computed in part (a), what will the value of entries in optimal weight vector $\hat{\theta}$ be?

   If we use stochastic gradient descent to iteratively approximate optimal θ, how will value of cross entropy loss change throughout the process?

4. Describe why logistic regression can be understood as a "generalized linear classification model". Where does "linear model" come from? Why can it create "nonlinear decision boundaries"?

Exercise 4

This question is to ensure a mathematical understanding of how the loss function leads to training a logistic regression model. Complete this section without referring to the notes.

1. Write down the cross entropy loss for logistic regression.
2. Now take the derivative of this loss with respect to the weights $\vec{\theta}$.

Exercise 5

Potpourri. Long answers aren't necessary for this part.

1. Is there a closed form solution to logistic regression like there is for linear regression?
2. What's the name of the encoding we use for multi-class classification?
3. What is the function that generalises the sigmoid to multi-class classifcation problems?