

50.007 Machine Learning, Fall 2021

Homework 3

Due: 3 November 2021

This homework is prepared by Prof. Berrak Sisman,
and will be graded by Ms. Zongyang Du (TA).

1 Logistic Regression

Question 1.1 [20 pts]

Please indicate whether the following statements are true (T) or false (F).

- a) Logistic regression is a supervised machine learning algorithm.
- b) Logistic regression is mainly used for regression, not classification.
- c) It possible to design a logistic regression using a Neural Network Algorithm.
- d) Logistic regression outputs a probability or confidence score, i.e., a value between 0 and 100.

Question 1.2 [20 pts]

Suppose that you have trained a logistic regression classifier, and it outputs on a new example a prediction $h_{\theta}(x) = 0.28$. This means (check all that apply):

- 1) Our estimate for $P(y = 0|x; \theta)$ is 0.28
- 2) Our estimate for $P(y = 0|x; \theta)$ is 0.72
- 3) Our estimate for $P(y = 1|x; \theta)$ is 0.28
- 4) Our estimate for $P(y = 1|x; \theta)$ is 0.72

Please explain your answer.

Question 1.3 [20 pts]

Suppose you train a logistic classifier $h_{\theta}(x) = g(\theta_0 + x_1\theta_1 + x_2\theta_2)$, and obtain $\theta = [3 \ -3 \ 0]^T$. Please formulate the decision boundary of your classifier. Note that this is a binary classification problem, which means class label y can be 0 or 1.

Question 1.4 [20 pts]

Suppose you train a logistic classifier $h_{\theta}(x) = g(\theta_0 + x_1\theta_1 + x_2\theta_2 + x_1^2\theta_3 + x_2^2\theta_4)$, and obtain $\theta = [-64 \ 0 \ 0 \ 1 \ 1]^T$. Please formulate the decision boundary of your classifier. Note that this is a binary classification problem, which means class label y can be 0 or 1.

Question 1.5 [20 pts]

In logistic regression, we find the parameters of a logistic (sigmoid) function that maximize the likelihood of a set of training examples. The likelihood is given as follows:

$$\prod_{i=1}^n P(y^{(i)}|x^{(i)}) \quad (1)$$

However, we re-express the problem of maximizing the likelihood as minimizing the following expression:

$$\frac{1}{n} \sum_{i=1}^n \log (1 + \exp (-y^{(i)} (\theta \cdot x^{(i)} + \theta_0))) \quad (2)$$

What is the benefit of optimizing the log-likelihood rather than the likelihood of the data? In other words, why is this expression computationally more “convenient”? (*Hint: try randomly generating, say, 1,000 probabilities in Python and multiplying them together as in Eq. 1.*)