Introduction to Machine Learning
Lab 9: Linear Determinant Analysis and Logistic Regression

Hongteng Xu

May 13, 2022

1 Motivation

- Implement the LDA training and testing methods.
- Implement the SGD of logistic regression and review the chain rule of gradients.
- Explore the unfairness issue hidden in the trained model.

2 Tasks

Please read Lecture 11 carefully before doing this lab work. In this Lab, we leverage the adult income dataset to train two classifiers (based on LDA and logistic regression, respectively). Each classifier takes people's profiles as input, and predicts whether they are high-income (e.g., $\geq$ 50k USD per year) or low-income (e.g., $<$ 50k USD per year).

1. **LDA:** Given a set of training and testing data, implement the training and testing methods of the linear determinant analysis model.
2. **Logistic regression:** Given a set of training and testing data, implement the training and testing methods of the logistic regression model. Here, the training method is the stochastic gradient descent (SGD).
3. **Gender fairness:** Given the learned classifiers, please find a method to check whether they are fair with respect to gender. You may find that **even if the gender information is not used in training, your classifiers may inherit the discrimination hidden in our society.... so sad:(** (Hint: Given the testing data, the fairness checker takes their prediction results and their gender information as input, and outputs the distribution of the results conditioned on different genders.)