Homework 01

CS3339 Machine learning (Enhanced) 2025 Fall*
School of Computer Science
Shanghai Jiao Tong University

Submission deadline: 23:59, October 26, 2025, Sunday

Submission to:

Please submit your homework in pdf/doc format to Canvas platform.

1 (30 points) Equivalence between LR and NB

Please prove that Naive Bayes (NB) and Logistic Regression are mathematically equivalent under certain conditions.

2 (70 points) Empirical investigations

Select at least two data sets from the link (Data A) below and then investigate classification performances of Logistic Regression (LR), Naive Bayes (NB), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and neural networks (e.g., MLP) on the selected data sets.

You may try different experimental settings, e.g., by varying the sample size of the training set, trying data sets with different dimensions, and other configurations that may affect performance in your mind. You may also try different kernels for SVM.

Links to the data sets:

- Data A: `https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/`

Links to the codes for your references:

- LR: `https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression`
- NB: `https://scikit-learn.org/stable/modules/naive_bayes.html`
- LDA: `https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html`
- SVM: `https://www.csie.ntu.edu.tw/~cjlin/libsvm/`
- SVM: `http://scikit-learn.org/stable/modules/svm.html`
- MLP: `http://scikit-learn.org/stable/modules/neural_networks_supervised.html`

*Liqing Zhang: zhang-lq@cs.sjtu.edu.cn; Shikui Tu: tushikui@sjtu.edu.cn