

Introduction to Data Science and AI: LAB 3

Tung Kieu

May 15, 2023

1. Requirements

- 1.1 **Integrated Development Environment (IDE):** PyCharm.
- 1.2 **Version Control:** Git and TortoiseGit.
- 1.3 **Compiler & Interpreter:** Python 3 (WinPython on Windows or Anaconda on Linux).
- 1.4 **Additional Libraries:** Pandas, NumPy, SciPy, Matplotlib, Sklearn, (and PyTorch).
- 1.5 **Data Sets:** Iris and Diabetes

2. Assignment

- 2.1 Load data set Diabetes (<https://archive-beta.ics.uci.edu/dataset/34/diabetes>) by using NumPy.
- 2.2 Describe the dataset.
- 2.3 Partition the dataset into the training and the testing set with a ratio of 80/20.
- 2.4 Train a linear regression model (using Sklearn) on the training set.
- 2.5 Use the trained model on the testing set and report the error.
- 2.6 Train a linear regression model (using Sklearn) on the training set with only one feature.
- 2.7 Use the trained model on the testing set and report the error.
- 2.8 Visualize the linear model.
- 2.9 (BONUS) Train a linear regression model (using manual inverse matrix) on the training set.
- 2.10 (BONUS) Use the trained model on the testing set and report the error.

- 2.11 Load data set IRIS (<https://archive-beta.ics.uci.edu/dataset/53/iris>) by using NumPy.
- 2.12 Describe the dataset.
- 2.13 Use only samples from two arbitrary classes in the dataset.
- 2.14 Partition the dataset into the training and the testing set with a ratio of 80/20.
- 2.15 Train a logistic regression model (using Sklearn) on the training set.
- 2.16 Use the trained model on the testing set and report the accuracy.
- 2.17 Use the full dataset.
- 2.18 Partition the dataset into the training and the testing set with a ratio of 80/20.
- 2.19 Train a k-NN model (using Sklearn) on the training set.
- 2.20 (BONUS) Train a k-NN model (using Sklearn) with different hyperparameters on the training set.
- 2.21 Use the trained model on the testing set and report the accuracy.
- 2.22 Train a naive Bayes model (using Sklearn) on the training set.
- 2.23 (BONUS) Train a naive Bayes model (using Sklearn) with different hyperparameters on the training set.
- 2.24 Use the trained model on the testing set and report the accuracy.
- 2.25 Train a decision tree model (using Sklearn) on the training set.
- 2.26 (BONUS) Train a decision tree model (using Sklearn) with different hyperparameters on the training set.
- 2.27 Use the trained model on the testing set and report the accuracy.

3. Submission

- 3.1 Write the report in LaTeX.
- 3.2 The report should have code, results, chart figures, and explanation.
- 3.3 The report should be submitted in one week after the lab date.