Source code description CS2750 ML project

Qingyang Li

April 27, 2022

1   Packages

The project is build on jupyter notebook, program language is Python. The packages mainly used in this project as followed:

- numpy (data array store and load)

- glob (for load images)

- seaborn (heatmap)

- matplotlib (graph)

- opencv-python (read images)

- tensorflow.keras (AlexNet)

- pickle (save and load models)

- slearn (ML models and data set split)

2   Run

Image set I used in this project is already selected by *img_selection.ipynb* and store in *data* directory due to the size limit. Origin dataset:https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images

2.1   data preprocess

In *code* directory, run data_preprocess.ipynb, the program will generate data array store in npy for training and testing purpose.

2.2   Training models

For model training, run the folloing code

- Run *LR.ipynb* can train the logistic regression model

- Run *KD.ipynb* can train the KNN model

- Run *SVM.ipynb* can train the SVM model

- Run *RandomForest.ipynb* can train the Random Forest model

- Run *AlexNet.ipynb* can train the AlexNet model

2.3   Result and Evaluation

Run *performance.ipynb*, execute the cells to test performance for different models (Becareful KNN model! It can take much longer time for testing!)