README

This is the illustrative document for exercise 3 of Machine Learning Course.

Content

Files in the package

1. *Exercise3_20190926.pdf* - The assignment file.
2. *run.py* - The running file.
3. *utils.py* - The util functions file.
4. *exercise_3_ML_季林成_2017012775_经71.pdf* - The experiment report file.
5. Given dataset files
    1. *train_10gene.csv*
    2. *train_10gene_sub.csv*
    3. *train_label.csv*
    4. *train_label_sub.csv*
    5. *test_10gene.csv*
    6. *test2_10gene.csv*
    7. *test_label.csv*
    8. *test2_label.csv*
6. Preprocessing dataset files
    1. *unquoted_(train_10gene/train_10gene_sub/train_lable/train_label_sub/test_10gene/test_label/test2_10gene/test2_label).csv*
    2. *transposed_unquoted_(train_10gene/train_10gene_sub/train_lable/train_label_sub/test_10gene/test_label/test2_10gene/test2_label).csv*
7. *(poly or linear)_train_1/2_test_1/2.log* - log files
8. README.md

Network structure

The networks applied in the package is based on scikit-learn.

Programming environment

Python 3.7.x(stable), sklearn(can be applied by pip)

How to run the code

1. open *run.py*
2. search for "_trainset_" to locate the naming code block in the middle
3. search for "train_data"(or train_target, test_data, test_target, core_function) to locate the main code block at the end
4. change the kernel function: replace "poly" in core_function with other kernels, such as "linear", "rbf", "sigmoid"
5. change the trainset/testset: replace corresponding address with addresses you like
6. run *run.py*
7. NOTE: when warning info "parell table not identical" appears, please run the code again