INT410: Advanced Pattern Recognition **Fall 2020**

Assignment 2: Non-parametric Classifiers

Lecturer: Non-parametric Classifiers *Unit: SAT of XJTLU*

Kaizhu Huang

Disclaimer:

1. Lab reports deadlines are strict. University late submission policy will be applied.

2. Collusion and plagiarism are absolutely forbidden (University policy will be applied).

2.1 Objectives

- Implement the Parzen window and KNN algorithm
- In this experiment, we will use the publicly dataset to verify our algorithm. Download the UCI Iris dataset: https://archive.ics.uci.edu/ml/datasets/Iris
- **Warnings:** For KNN and Parzen algorithms, there is no training stage.

2.2 Estimation of Classification Methods

- Read the Iris dataset into a list and shuffle it with the **random.shuffle** method. Hint: fix the random seed (e.g. **random.seed(17)**) before calling **random.shuffle**
- Split the dataset as five parts to do cross-fold validation: Each of 5 subsets was used as test set and the remaining data was used for training. The 5 subsets were used for testing rotationally for evaluating the classification accuracy.

2.3 Parzen Window Method (50 marks)

- Separate the training dataset into two groups by their labels.
- Estimate the prior class probability $P(\omega_k)$

$$P(\omega_k) = \frac{n_k}{\sum_{i=1}^{K} n_i}$$

where n_k is the number of examples from the class ω_k.

- For any test example x, the conditional probability $P(x|\omega_k)$ are computed as

$$P(x|\omega_k) = \frac{1}{n_k} \sum_{x_i \in \omega_k} \frac{1}{h^d} \phi\left(\frac{x - x_i}{h}\right)$$

where h is hyperparameter (or user-defined parameter) of model and $\phi(u)$ is defined as

$$\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

- According the Bayesian rule, we know that

  $$P(\omega_k|x) \propto P(\omega_k)P(x|\omega_k)$$

  The example x is assigned to the label with the maximum $P(\omega_k|x)$.
- Show the figure that the accuracy performance changes with the hyperparameter h (real number).

2.4 KNN Method (50 makrs)

- For each test example, we find K (K is a hyperparameter of KNN) nearest neighbor examples from n labeled training examples, where the distance is measured by the Euclidean distance (with the norm $\|\cdot\|_2$).
- Estimate the posterior probability $P(\omega_k|x)$ as

  $$P(\omega_k|x) = \frac{K_i}{K}$$

  where K_i is the number of the examples from K nearest neighbor examples.
- Decision rule: The example x is assigned to the label with the maximum $P(\omega_k|x)$
- Show the figure that the accuracy performance changes with the hyperparameter K (natural number).

2.5 Lab Report

- Write a short report which should contain a concise description of your results and observations.
- **Please insert the clipped running image into your report for each step.**
- Submit the report and the source codes electronically into LearningMall.
- The report is encouraged to be written with the latex typesetting language.
- The report in pdf format and source codes of your implementation should be zipped into a single file. The naming of report is as follows:

  e.g. StudentID_LastName_FirstName_LabNumber.zip (123456789_Einstein_Albert_1.zip)

2.6 Hints

Please refer to the lecture slides and PRML book (Section 4.1.6) for more details.

- Latex IDE: texstudio
- Python IDE: pycharm
- Use the python numpy library flexibly.