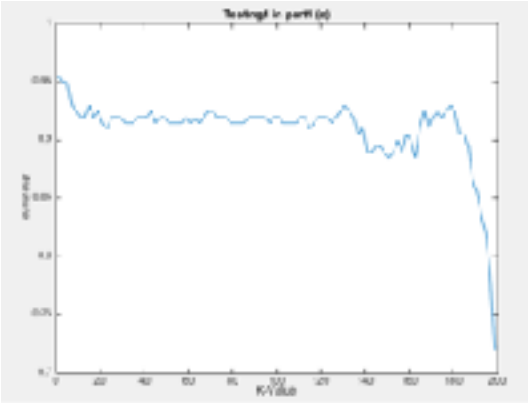
COMP24111 EX 2 Report

Boyin Yang 10071239

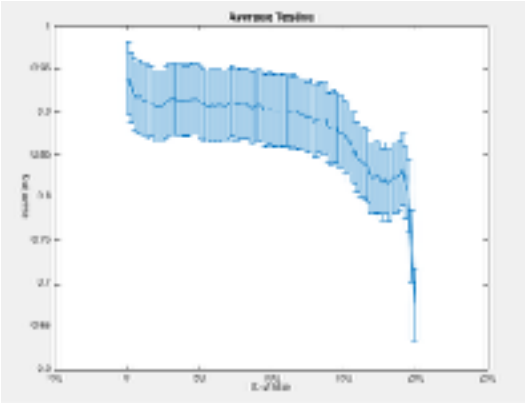
Oct 23 2016

Part 1 Training and Testing

a) Using *shufflerows()* function to get original data in *data.* Then using *horzcat()* and *vercat()* to merge the data and its labels.
b) K = 1 means only finding 1 training data that is closest to the testing data, which will also define the testing data. When K gets bigger, more training data could be used to define the testing data, whose value will be defined by the larger number of training data.
c) Pic.1 is the accuracy of different K value. Training is helping a machine to set a standard of two different data. While testing is using training data to define the testing data. Pic.1 is the accuracy of varying K value.



Pic. 1



Pic. 2

d) Repeat the process for (c) many tomes will not get the same behaviour, because the testing data is different. Pic. 2 is the average accuracy, and standard deviation as error bars.

Part 2 Feature Extractors for Digit

First, reduce the unimportant data from dataset by setting every pixel less than 100 to 0.
Second, 2 dimension to 1 dimension.
Pic. 3 is the result of accuracy as feature extractors.

Pic. 3

Part 3 Perceptrons

Training time is less than 0.3s.
When training interaction time is big enough (eg. more than 50), accuracy could be 100%.

Part 4 Logistic Regression

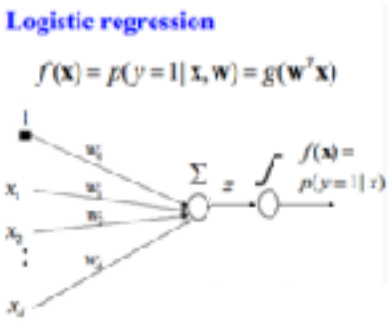
As I learned in lecture, perceptron is a linear regression, which may get some big activations. What logistic regression does is to diminish the activations to a range of 0~1. That is the probability of an event happening.

In our 2-class KNN example, in order to judge the handwriting number is 3 or 8, we can get the probability of the testing data. If the probability of being 3 is bigger than 50%, then that could be 3, otherwise it could be 8.

But in a lot of cases, things cannot be defined totally true or wrong. So perceptrons may not be the best choice. Logistic regression can classify things by the probabilities of happening.

Pic. 4 and Pic. 5 are the conception diagram of linear regression and logistic regression.

Pic. 4



Pic. 5