Machine Learning Project 1

Boyi Zheng(G43492924)

Resources: Python.

Packages references and usage: pandas(dataFram), numpy(basic algebra), itertools(basic for counter), time/datetime(time counting and print), os, collections-Counter(a efficient way of counting labels), matplotlib(histogram ), sklearn(only for confusion matrix).

Part 1: Pima

description: Based on the existing medical information, the probability of the onset of diabetes in the Nepima Indians in 5 years is predicted.

Columns and details:

0 pregnants: Number of pregnancies

1 glucose: Plasma glucose concentration after 2 hours in oral glucose tolerance test, shows the Diabetes degree

2 blood\_pressure: A health indicator

3 Triceps\_skin\_fold\_thickness: Triceps skinfold thickness, Determine the degree of obesity

4 serum\_insulin: Postprandial serum insulin , Diabetes measures

5 BMI: now still using BMI as a measure of health related to weight and height

6 Diabetes\_pedigree\_function:Genetically related

7 age apparently, years that breathing

8 outcome of claasification

|  |  |
| --- | --- |
| 图表, 直方图  描述已自动生成 | 图表, 直方图  描述已自动生成 |
| 图表, 直方图  描述已自动生成 | 图表, 直方图  描述已自动生成 |
| 图表, 直方图  描述已自动生成 | 图表, 直方图  描述已自动生成 |
| 图表, 直方图  描述已自动生成 | 图表, 直方图  描述已自动生成 |

表格

描述已自动生成

The comparison of origin outcome and model prediction outcome

|  |  |
| --- | --- |
| 图表, 直方图  描述已自动生成 | 图表, 直方图  描述已自动生成 |

Part 1: MNIST

The famous MNIST dataset. Almost every machine learning initiate with this dataset.

Each image is 28 pixels in height and 28 pixels in width, for a total of 784 pixels in total. Each pixel has a single pixel-value associated with it, indicating the lightness or darkness of that pixel, with higher numbers meaning darker. This pixel-value is an integer between 0 and 255, inclusive.

The training data set, (train.csv), has 785 columns. The first column, called "label", is the digit that was drawn by the user. The rest of the columns contain the pixel-values of the associated image.

Each pixel column in the training set has a name like pixelx, where x is an integer between 0 and 783, inclusive. To locate this pixel on the image, suppose that we have decomposed x as x = i \* 28 + j, where i and j are integers between 0 and 27, inclusive. Then pixelx is located on row i and column j of a 28 x 28 matrix, (indexing by zero).

表格

描述已自动生成

图片包含 表格

描述已自动生成

Runtime:

For Pima: 1sec

For MNIST: pick 200 for train and K=3 20min21sec in total.

All runtime based on single processer.