**ENGN2520 Homework 3**

**Ming Xu (Banner ID: B01532164)**

**Problem1**

***(a)***

We know that: , where

is the priori distribution of ***w*** and . In this case:

is the likelihood. In this case:

Combine the equations above:

Let be the right part the equation above, in this case:

The MAP estimate of ***w*** is the vector maximizing the posterior probability of ***w*** given ***T***:

Denote as the log of , then:

Ignore the constant times to get :

Therefore,

Let , then can be written as:

***(b)***

**Problem2**

Let be the ***ith*** training data from the data set ***T***. The likelihood can be written as:

Note that, from a data , there are n features, namely .

Also, distributed according to a Bernoulli distribution with mean , then:

Therefore,

Denote as the log of , then:

Taking the derivative of this expression with respect to , we get:

Let , which is the total number of ones in the ***jth*** feature of all the data in the training set.

**Problem3**

***(a)***

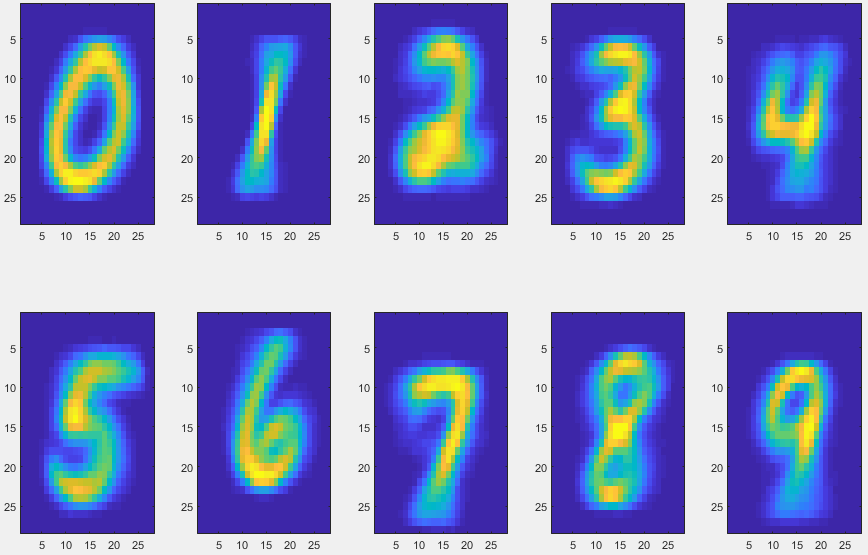
The result is very similar the result of Problem2.

The maximum likelihood estimate for parameters is:

Where n is the number of training data of class and is the total number of ones in the pixel of the training set.

***(b)***

The visualization of the models is shown as below:



***(c)***

The total number of testing data that is correctly classified is 3942, so the fraction is 78.84%.

The confusion matrix is show as below:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **0** | 433 | 0 | 4 | 0 | 2 | 33 | 13 | 0 | 14 | 1 |
| **1** | 0 | 470 | 2 | 3 | 0 | 14 | 4 | 0 | 7 | 0 |
| **2** | 16 | 12 | 370 | 36 | 8 | 2 | 6 | 11 | 32 | 7 |
| **3** | 1 | 10 | 4 | 414 | 6 | 23 | 6 | 13 | 10 | 13 |
| **4** | 4 | 1 | 11 | 0 | 360 | 6 | 13 | 4 | 7 | 94 |
| **5** | 19 | 1 | 3 | 66 | 19 | 344 | 7 | 5 | 16 | 20 |
| **6** | 13 | 8 | 30 | 0 | 6 | 33 | 407 | 0 | 3 | 0 |
| **7** | 8 | 16 | 8 | 4 | 15 | 0 | 2 | 389 | 9 | 49 |
| **8** | 4 | 16 | 13 | 50 | 13 | 18 | 1 | 4 | 346 | 35 |
| **9** | 3 | 7 | 4 | 8 | 49 | 7 | 0 | 7 | 6 | 409 |

***(d)*** Source code

**1.** **Function to train the classification model: “trainModelForDigit.m”**

function [model] = trainModelForDigit(trainSet)

[rowNum, ~] = size(trainSet);

model = sum(trainSet)/rowNum;

end

**2. Function to calculate the probability of the given data and given class:” calculateProbability.m”**

function [probability] = calculateProbability(data,model)

result1 = model.^data;

model = 1-model;

data = 1-data;

result2 = model.^data;

result = result1.\*result2;

probability = prod(result);

end

**3. Function to classify the testing set by models trained on the training set:” classifyDigit.m”**

function [classifyResult] = classifyDigit(testData,models)

%initialization

[dataCount,~] = size(testData);

result = zeros(dataCount,10);

classifyResult = zeros(1,10);

%loop all testing data

for row = 1:dataCount

%loop each class to calculate probability

for col = 1:10

result(row,col) = calculateProbability(testData(row,:),models(col,:));

end

%find the max value and index of the probability

[~, index] = max(result(row,:));

%classify

classifyResult(1,index) = classifyResult(1,index) + 1;

end

end

**4. Main function:”main.m”**

load 'digits';

models = zeros(10,784);

%% Train model for each digit

for i = 1:10

traindata = sprintf('%s%d','train',i-1);

models(i,:) = trainModelForDigit(eval(traindata));

end

%% Draw the visualization of the models

for i = 1:10

subplot(2,5,i);

imagesc(reshape(models(i,:),28,28)');

end

%% Classify each testing set

confusion = zeros(10,10);

correct = 0;

for i = 1:10

testdata = sprintf('%s%d','test',i-1);

confusion(i,:) = classifyDigit(eval(testdata),models);

correct = correct + confusion(i,i);

end