Name: Jingyi Yuan
UNI: jy2736
Class: EECS E6892

Homework 01

Problem 1

Suppose the case that the chosen door contains a prize is Case A.
P(A) = 1/3
The case that the gameshow host opens a door without a prize is Case B.
Since the host has already know where the prize is, P(B) = 1

$$\mathrm{P(A|B)} = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)}$$

If A happens, then the probability of the host to choose a door without a prize is 1, that is, $P(B|A) = 1$, so

$$\mathrm{P(A|B)} = \frac{P(A)P(B|A)}{P(B)} = \frac{\frac{1}{3} * 1}{1} = \frac{1}{3}$$

Thus she should switch the door.

Problem 2

Problem 2.

Since $X_i \sim$ Multinomial (π).

$$P(x_1=\pi_1, x_2=\pi_2 \dots, x_k=\pi_k) = \frac{n!}{x_1! x_2! \cdots x_k!} \pi_1^{x_1} \pi_2^{x_2} \dots \pi_k^{x_k}.$$

$$= p(x_1, \dots x_k | \pi_1 \dots \pi_k) \Rightarrow \underline{\text{likelihood.}} \quad \sum \pi_i = 1 ,$$

We try Dirichlet distribution as the prior for π:

$$p(\pi_1 \dots \pi_k) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} \pi_i^{\alpha_i - 1} , \quad \sum_{i=1}^{k} \pi_i = 1 \Rightarrow \underline{\text{prior}}$$

$$\therefore p(\pi_1 \dots \pi_k | x_1 \dots x_k) \propto p(x_1 \dots x_k | \pi_1 \dots \pi_k) \cdot p(\pi_1 \dots \pi_k).$$

$$\propto \frac{n!}{x_1! x_2! \dots x_k!} \pi_1^{x_1} \pi_2^{x_2} \dots \pi_k^{x_k} \cdot \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} \pi_i^{\alpha_i - 1}$$

$$\propto \prod_{i=1}^{k} \pi_i^{x_i} \cdot \prod_{i=1}^{k} \pi_i^{\alpha_i - 1}$$

$$\propto \prod_{i=1}^{k} \pi_i^{x_i + \alpha_i - 1} \qquad \Rightarrow \underline{\text{posterior}}$$

so, the posterior distribution is also a Dirichlet distribution.

$\therefore$ the conjugate prior for π is the Dirichlet distribution. The parameters of the posterior distribution are product of some gamma functions, and have no relationship with π_i.

Problem 3

a) $p(\mu,\lambda) \propto p(\mu|\lambda)p(\lambda)$

$$\propto \lambda^{\frac{1}{2}} \exp\left(-\frac{\lambda}{2a}\mu^2\right) \lambda^{b-1} e^{-\lambda c}$$

$$\propto \lambda^{b-\frac{1}{2}} \exp\{-c\lambda\} \exp\left\{-\frac{\lambda\mu^2}{2a}\right\}$$

$\therefore \mu, \lambda \sim$ Normal Gamma $(0, a^{-1}, b, c)$.

Posterior :

$$p(\mu,\lambda|x) \propto p(x|\mu,\lambda)p(\mu,\lambda)$$

$$\propto \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}} \lambda^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\lambda(x_i-\mu)^2\right\}\right) \lambda^{b-\frac{1}{2}} \exp\{-c\lambda\} \exp\left\{-\frac{\lambda\mu^2}{2a}\right\}$$

$$\propto \lambda^{\frac{n}{2}} \cdot \exp\left\{-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \lambda^{b-\frac{1}{2}} \exp\{-c\lambda\} \exp\left\{-\frac{\lambda\mu^2}{2a}\right\}.$$

since: $\exp\left\{\sum_{i=1}^{n}(x_i-\mu)^2\right\} \propto \exp\left\{\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2\right\}$ $\qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$\propto \exp\left\{\sum_{i=1}^{n}\left[(x_i-\bar{x})^2+(\bar{x}-\mu)^2\right]\right\}$$

$$\propto \exp\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2\right\}$$

$$\therefore p(\mu,\lambda|x) \propto \lambda^{\frac{n}{2}+b-\frac{1}{2}} \exp\left\{-\lambda\left[\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2+c\right]\right\} \exp\left\{-\frac{\lambda}{2}\left(\underline{\frac{1}{a}\mu^2+n(\bar{x}-\mu)^2}\right)\right\}$$

since. $\underline{\frac{1}{a}\mu^2+n(\bar{x}-\mu)^2} = \frac{1}{a}\mu^2+n\mu^2-2n\bar{x}\mu+n\bar{x}^2$

$$=\left(\frac{1}{a}+n\right)\mu^2-2n\bar{x}\mu+n\bar{x}^2$$

$$=\left(\frac{1}{a}+n\right)\left(\mu^2-\frac{2n\bar{x}}{1/a+n}\mu\right)+n\bar{x}^2$$

$$=\left(\frac{1}{a}+n\right)\left(\mu-\frac{n\bar{x}}{1/a+n}\right)^2+n\bar{x}^2-\frac{(n\bar{x})^2}{1/a+n}$$

$$=\left(\frac{1}{a}+n\right)\left(\mu-\frac{n\bar{x}}{1/a+n}\right)^2+\frac{\frac{1}{a}n\bar{x}^2}{1/a+n}$$

$$\therefore p(\mu,\lambda|x) \propto \lambda^{\frac{n}{2}+b-\frac{1}{2}} \exp\left\{-\lambda\left[\underbrace{\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2+c+\frac{\frac{1}{a}n\bar{x}^2}{2(1/a+n)}}\right]\right\} \exp\left\{-\frac{\lambda}{2}\left(\underline{\frac{1}{a}+n}\right)\left(\mu-\frac{n\bar{x}}{1/a+n}\right)^2\right\}$$

$\therefore b_n = b+\frac{n}{2}$.

$C_n = \frac{1}{2}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{\frac{1}{a}n\bar{x}^2}{1/a+n}\right)+c$.

$\mu_n = \frac{n\bar{x}}{1/a+n}$

$a_n = \frac{1}{a}+n$.

$\therefore p(\mu,\lambda|x) =$ Normal Gamma (μ_n, a_n, b_n, C_n).

$\because p(\mu,\lambda|x) \propto p(\mu|\lambda,x)\cdot p(\lambda|x)$

$\therefore p(\mu|\lambda,x) \propto \lambda^{\frac{1}{2}} \exp\left\{-\frac{\lambda}{2}\left(\frac{1}{a}+n\right)\left(\mu-\frac{n\bar{x}}{a^{-1}+n}\right)^2\right\} \propto$ Normal (μ_n, a_n)

and $p(\lambda|x) \propto \lambda^{\frac{n}{2}+b-1} \exp\left\{-\lambda\left[\frac{1}{2}\sum_{i=1}^{n}(x_i-x)^2+c+\frac{\frac{1}{a}n\bar{x}^2}{2(a^{-1}+n)}\right]\right\}$

$\propto$ Gamma (b_n, C_n)

$\therefore$ $p(\mu|\lambda, x)$ and $p(\lambda|x)$ is another Normal - Gamma distribution while:

$$p(\mu|\lambda, x) = \text{Normal}\left(\frac{n\bar{x}}{a^{-1}+n},\ a^{-1}+n \right)$$

$$p(\lambda|x) = \text{Gamma}\left(b+\frac{n}{2},\ c+\left(\frac{1}{2} \sum_{i=1}^{n} (x_i-\bar{x})^2 + \frac{a^{-1} n \bar{x}^2}{a^{-1}+n} \right) \right).$$

b). $p(x^*|x_1 \dots x_n) = \int_0^\infty \int_{-\infty}^{\infty} p(x^*|\mu,\lambda)\, p(\mu,\lambda|x_1,\dots x_N)\, d\mu\, d\lambda$

$$= \int_0^\infty \int_{-\infty}^{\infty} \text{Normal}(\mu, \lambda^{-1})\ \text{Normal Gamma}(\mu_n, a_n, b_n, c_n)\, d\mu\, d\lambda$$

$$\propto \int_0^\infty \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \lambda^{-\frac{1}{2}} \exp\left\{-\frac{\lambda}{2}(x^*-\mu)^2\right\} \lambda^{b_n-\frac{1}{2}} \exp\{c_n\lambda\} \exp\left\{-\frac{\lambda}{2} a_n (\mu-\mu_n)^2\right\} d\mu\, d\lambda.$$

$$\propto \int_0^\infty \int_{-\infty}^{\infty} \lambda^{b_n-1} \exp\{-c_n\lambda\} \cdot \exp\left\{-\frac{\lambda}{2}\left[(x^*-\mu)^2 + a_n(\mu-\mu_n)^2 \right]\right\} d\mu\, d\lambda$$

$$\propto \int_0^\infty \int_{-\infty}^{\infty} \lambda^{b_n-1} \exp\{-c_n\lambda\} \cdot \exp\left\{-\frac{\lambda}{2}\left[(a_n+1)\mu^2 - 2(x^*+a_n\mu_n)\mu + x^{*2} + a_n\mu_n^2 \right]\right\} d\mu\, d\lambda.$$

sin $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \lambda^{-\frac{1}{2}} \exp\left\{-\frac{\lambda}{2}(x-\mu)^2\right\} dx = 1$

$\therefore \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \lambda^{-\frac{1}{2}} \exp\left\{-\frac{\lambda}{2}(\mu-x^*)^2\right\} d\mu = 1$

$$\therefore p(x^*|x_1 \dots x_n) \propto \int_0^\infty \lambda^{b_n-\frac{1}{2}} \exp\{-c_n\lambda\} \cdot \exp\left\{-\frac{\lambda}{2}\left[\frac{(x^*+a_n\mu_n)^2}{a_n+1} - (a_n\mu_n^2 + x^{*2})\right]\right\}$$

$$\propto \Gamma\left(b_n+\frac{1}{2}\right) \cdot \left\{-\frac{1}{2}\left[\frac{(x^*+a_n\mu_n)^2}{a_n+1} - (a_n\mu_n^2 + x^{*2} + 2c_n)\right]\right\}^{-b_n-\frac{1}{2}}.$$

$$\propto \Gamma\left(b_n+\frac{1}{2}\right) \left\{-\frac{1}{2} \cdot \frac{-a_n(x^*-\mu_n)^2 - 2c_n(a_n+1)}{a_n+1}\right\}^{-b_n-\frac{1}{2}}$$

$$\propto \Gamma\left(b_n+\frac{1}{2}\right) \left\{\frac{a_n(x^*-\mu_n)^2}{2(a_n+1)} + c_n\right\}^{-b_n-\frac{1}{2}}$$

since $b_n \cdot c_n$ are constant when n is certain.

$$p(x^*|x_1 \dots x_n) \propto \frac{\Gamma(b_n+\frac{1}{2})}{\Gamma(b_n)\sqrt{\pi(a_n+1)c_n}} \left(\frac{a_n(x^*-\mu_n)^2}{2(a_n+1)c_n} + 1 \right)^{-b_n-\frac{1}{2}}$$

for Students t distribution: $f_v(\mu, t) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\,\Gamma(\frac{v}{2})} \left(1+\frac{(t-\mu)^2}{v}\right)^{-\frac{v+1}{2}}$

$\therefore v = 2b_n,\quad \mu = \mu_n.$

$\therefore p(x^*|x_1 \dots x_n) \propto t_{2b_n}\left(x^* \middle| \mu_n, \frac{c_n(a_n+1)}{b_n a_n}\right).$

$\therefore$ the predictive distribution on a new observation x^* is a Student's t distribution

Problem 4

a)

```
clear;
close all

load mnist_mat
[row_Xtrain,column_Xtrain] = size(Xtrain);
[row_Xtest,column_Xtest] = size(Xtest);
number1 = sum(ytrain);
number0 = column_Xtrain - number1;
Sample1 = Xtrain(:,1+number0:column_Xtrain);
Sample0 = Xtrain(:,1:number0);
%calculate the means and variances of x
u1=mean(Sample1')';
u0=mean(Sample0')';
sigm1=var(Sample1')';
sigm0=var(Sample0')';

%calculate the means and variances of t distribution
a = 1;
b = 1;
c = 1;
e = 1;
f = 1;
un1 = (number1*u1)/(a+number1);
un0 = (number0*u0)/(a+number0);
an1 = 1 + number1;
an0 = 0 + number0;
bn1 = b + number1/2;
bn0 = b + number0/2;
cn1 = c + number1*sigm1/2 + a*number1*(u1 .* u1)/(2*(a + number1));
cn0 = c + number0*sigm0/2 + a*number0*(u0 .* u0)/(2*(a + number0));
```

```
freedom1 = 2*bn1;
freedom0 = 2*bn0;
sigma1 = sqrt(cn1*(an1 + 1)/(bn1*an1));
sigma0 = sqrt(cn0*(an0 + 1)/(bn0*an0));

%normalization
X1minus = bsxfun(@minus,Xtest',un1')';
X0minus = bsxfun(@minus,Xtest',un0')';
X1 = bsxfun(@times,X1minus,(1./sigma1));
X0 = bsxfun(@times,X0minus,(1./sigma0));
psum1 = 1;
psum0 = 1;

%Posterior predictive
for i = 1:15
    p_x_yequal1 = tpdf(X1(i,:),freedom1);
    p_x_yequal0 = tpdf(X0(i,:),freedom0);
    psum1 = psum1 .* p_x_yequal1;
    psum0 = psum0 .* p_x_yequal0;
end
p_yequal1_y = (e + number1)/(column_Xtrain + e + f);
p_yequal0_y = (f + number0)/(column_Xtrain + e + f);
px1 = psum1 * p_yequal1_y;
px0 = psum0 * p_yequal0_y;
p = zeros(1,column_Xtest);
x_wrong = zeros(1,column_Xtest);

```

```matlab
%QUESTION b, confusion matrix
confsMatrix = zeros(2,2);
number1iny = sum(ytest);
number0iny = column_Xtest - number1iny;
for i = 1:column_Xtest
    if (px0(i)>px1(i))
        p(i) = 0;
    else p(i) = 1;
    end
    if (i<number0iny+1)
        if (p(i) == ytest(i))
            confsMatrix(1,1) = confsMatrix(1,1) + 1;
        else
            confsMatrix(1,2) = confsMatrix(1,2) + 1;%4's classified as 9's
            x_wrong(i)= 1;
        end
    else
        if (p(i) == ytest(i))
            confsMatrix(2,2) = confsMatrix(2,2) + 1;
        else
            confsMatrix(2,1) = confsMatrix(2,1) + 1;%9's classified as 4's
            x_wrong(i)= 1;
        end
    end
end
rightNum = confsMatrix(1,1) + confsMatrix(2,2);
wrongNum = confsMatrix(1,2) + confsMatrix(2,1);
predicProb0 =  px0.*(1./(px0 + px1));
predicProb1 =  px1.*(1./(px0 + px1));

%a
correctness = rightNum/column_Xtest;

%c, show three misclassifed images
misImg = zeros(15,3);
j = 1;
index_mis = zeros(1,3);%records the positions of the misclassified numbers in Xtest
for i = 1:column_Xtest
    if(x_wrong(i) == 1)
        misImg(:,j) = Xtest(:,i);
        index_mis(j) = i;
        j = j+1;
        if (j == 4)
            break;
        end
    end
end

for i = 1:3
    x2 = Q * misImg(:,i);
    x_show = reshape(x2,28,28);
    subplot(2,3,i)
    imshow(x_show);
    fprintf('The predictive probility equals 0 of the %d misclassified is = %d\n',i,predicProb0(index_mis(i)));
    fprintf('The predictive probility equals 1 of the %d misclassified is = %d\n',i,predicProb1(index_mis(i)));
end

%d, three most ambiguous predictions
prdicProb = abs(predicProb0 - predicProb1);
[ambPred, index] = sort(prdicProb,'ascend');
for i = 1:3
    x3 = Q * Xtest(:,index(i));
    x_show3 = reshape(x3,28,28);
    subplot(2,3,i+3)
    imshow(x_show3);
    fprintf('The predictive probility equals 0 of the %d ambiguous number is = %d\n',i,predicProb0(index(i)));
    fprintf('The predictive probility equals 1 of the %d ambiguous number is = %d\n',i,predicProb1(index(i)));
end
```

The accuracy of the classifier is:

```
correctness =

    0.9141
```

b)

The confusion matrix is shown below. As it is suggested in the comments:
confsMatrix(1,1) = 956 is the total number of 4's classified as 4's.
confsMatrix(1,2) = 26 is the total number of 4's classified as 9's.
confsMatrix(2,1) = 145 is the total number of 9's classified as 4's.
confsMatrix(2,2) = 864 is the total number of 9's classified as 9's.

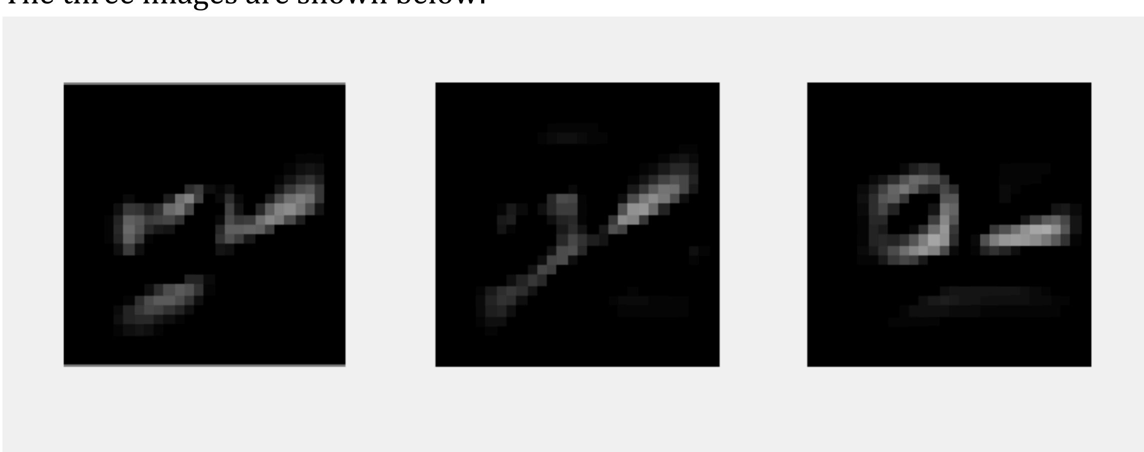
```
confsMatrix =

   956    26
   145   864
```

c)

The predictive probabilities of the three misclassified digits are shown below:

```
The predictive probility equals 0 of the 1 misclassified is = 3.752144e-01
The predictive probility equals 1 of the 1 misclassified is = 6.247856e-01
The predictive probility equals 0 of the 2 misclassified is = 3.856709e-01
The predictive probility equals 1 of the 2 misclassified is = 6.143291e-01
The predictive probility equals 0 of the 3 misclassified is = 4.148132e-01
The predictive probility equals 1 of the 3 misclassified is = 5.851868e-01
```

The three images are shown below:



d)

The predictive probabilities of the three most ambiguous predictions are shown below:

```
The predictive probility equals 0 of the 1 ambiguous number is = 5.001732e-01
The predictive probility equals 1 of the 1 ambiguous number is = 4.998268e-01
The predictive probility equals 0 of the 2 ambiguous number is = 4.991756e-01
The predictive probility equals 1 of the 2 ambiguous number is = 5.008244e-01
The predictive probility equals 0 of the 3 ambiguous number is = 5.015105e-01
The predictive probility equals 1 of the 3 ambiguous number is = 4.984895e-01
```

The three images are shown below: