EECS 6892
BMML HW1

UNI: sl3763
Name: Sung-Yen Liu

Problem 1.

Define A = number of door the host opens
B = number of door the prize behind
C = number of door the player picks

First, we consider some definition and constraints,
① $p[B=1] = p[B=2] = p[B=3] = \frac{1}{3}$

② $P[A=B] = \sum_{b=1}^{3} P[A=B \mid B=b] P[B=b] = 0$ (host never oper the door with prize)

③ $P[A=C] = 0$ (host never open the door the player picks)

Now we consider the game,
① The player makes a choice X $\Rightarrow P[C=B] = \sum_{b=1}^{3} P[C=B \mid B=b] P[B=b] = \frac{1}{3}$
② The host opens a door (neither the one player picks nor the prize behind)
③ Prob. of door of the prize and door of player picks are different:
(A≠B, B≠C, C≠A)

$$\begin{aligned} p[C \neq B] = & P[A=1, B=2, C=3] + P[A=1, B=3, C=2] \\ & + P[A=2, B=1, C=3] + P[A=2, B=3, C=1] \\ & + P[A=3, B=1, C=2] + P[A=3, B=2, C=1] = \frac{1}{9} \times 6 = \frac{2}{3} > \boxed{\text{if stay, prob} = \frac{1}{2}} \end{aligned}$$

By conclusion, the player should always change the selection to get better chance to win the prize under the condition that the host opens a door.

Problem 2.

For a single vector X with $\pi = (\pi_1, \dots, \pi_K)$,
we can express the likelihood as $p(x \mid \pi) = \prod_{k=1}^{K} \pi_k^{x_k}$

Now for a dataset with N i.i.d. vectors,

$$p(D \mid \pi) = p(X_1, \dots, X_n \mid \pi) \overset{\text{iid}}{=} \prod_{i=1}^{n} p(X_i \mid \pi) = \prod_{i=1}^{n} \prod_{k=1}^{K} \pi_k^{x_{ik}} = \prod_{k=1}^{K} \pi_k^{m_k}$$

where $m_k = \sum_{i=1}^{n} x_{ik}$

since we are finding conjugate prior, we can guess some features of distribution:

① the prior $p(\pi \mid \alpha) \propto \prod_{k=1}^{K} \pi_k^{\alpha_k - 1}$ (similar to the form of likelihood)

② Inspecting beta-distribution, we may assume $p(\pi \mid \alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_K)} \prod_{i=1}^{K} \pi_k^{\alpha_k - 1}$

This is called Dirichlet distribution which is the conjugate prior of multinomial distribution. $Dir(\pi|\alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)} \prod_{k=1}^{K} \pi_k^{\alpha_k - 1}$, where $\alpha = (\alpha_1, \dots, \alpha_K)^T$ and $\sum_{k=1}^{K} \alpha_k = \alpha_0$

$$p(\pi|D) \propto p(D|\pi)\, p(\pi)$$

$$\propto \left(\prod_{k=1}^{K} \pi_k^{m_k}\right) \times \left(\prod_{k=1}^{K} \pi_k^{\alpha_k - 1}\right) = \prod_{k=1}^{K} \pi_k^{m_k + \alpha_k - 1}$$

Following Bayes rule, the posterior is distributed as $Dirichlet(\pi | \alpha + m)$ where m is a K-dimension vector and $m_i = \sum x_i$

The feature of posterior's parameters : $\alpha' = \alpha + \underbrace{m}_{\text{the count of experiments}}$

Problem 3. (a)

Likelihood:

$$p(D|\mu,\lambda) \overset{\text{problem def.}}{=\!=} N(D|\mu,\lambda^{-1}) = \prod_{i=1}^{N} \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda}{2}(x_i-\mu)^2\right\}$$

$$= \left(\frac{\lambda}{2\pi}\right)^{\frac{N}{2}} \cdot \exp\left\{-\frac{\lambda}{2}\cdot\sum_i^N (x_i-\mu)^2\right\} = \left(\frac{\lambda}{2\pi}\right)^{\frac{N}{2}} \cdot \exp\left\{-\frac{\lambda}{2}\left(\sum_i x_i^2 - 2\mu\sum_i x_i + N\mu^2\right)\right\}$$

$$= \left(\frac{\lambda}{2\pi}\right)^{\frac{N}{2}} \exp\left\{-\frac{\lambda}{2}\left[N(\mu-\bar{x})^2 + \sum_{i=1}^{N}(x_i-\bar{x})^2\right]\right\} \qquad \left(\bar{x} = \frac{1}{n}\sum x_i\right)$$

Prior:

$$p(\mu,\lambda) = p(\mu|\lambda)\cdot p(\lambda) \overset{\text{problem def.}}{=\!=} N(\mu|0, a\lambda^{-1}) \cdot \text{Gamma}(b,c)$$

With unknown both mean and precision, this is the conjugate prior of the given Gaussian distribution, identified as Normal-Gamma distribution.

$$\text{NormalGamma}(\mu,\lambda|\mu_0,\kappa_0,\alpha_0,\beta_0) \overset{\text{def}}{=} N(\mu|\mu_0,(\kappa_0\lambda)^{-1})\,\text{Gam}(\lambda|\alpha_0,\beta_0)$$

$$= \frac{1}{Z_{NG}(\mu_0,\kappa_0,\alpha_0,\beta_0)} \lambda^{\frac{1}{2}} \exp\left(-\frac{\kappa_0\lambda}{2}(\mu-\mu_0)^2\right)\lambda^{\alpha_0-1}e^{-\lambda\beta_0}$$

$$= \frac{1}{Z_{NG}} \lambda^{\alpha_0-\frac{1}{2}} \exp\left(-\frac{\lambda}{2}\left[\kappa_0(\mu-\mu_0)^2+2\beta_0\right]\right)$$

$$\text{where } Z_{NG}(\mu_0,\kappa_0,\alpha_0,\beta_0) = \frac{\Gamma(\alpha_0)}{\beta_0^{\alpha_0}}\left(\frac{2\pi}{\kappa_0}\right)^{\frac{1}{2}}$$

As the problem gives $NG(\mu,\lambda|\underline{0, a^{-1}, b, c})$

$$Z_{NG} = \frac{\Gamma(b)}{c^b}\cdot(2\pi a)^{\frac{1}{2}}$$

and $p(\mu,\lambda) = \frac{1}{Z_{NG}}\lambda^{b-\frac{1}{2}}\exp\left(-\frac{\lambda}{2}\left[\frac{\mu^2}{a}+2c\right]\right)$

Posterior:

As we know it's conjugate prior, we can also expect posterior to be Normal-Gamma distribution.

General form:

$$p(\mu,\lambda|D) \propto NG(\mu,\lambda|\mu_0,\kappa_0,\alpha_0,\beta_0)\,p(D|\mu,\lambda)$$

$$\propto \lambda^{\frac{1}{2}} e^{-\frac{\kappa_0\lambda(\mu-\mu_0)^2}{2}} \lambda^{\alpha_0-1} e^{-\beta_0\lambda} \times \lambda^{\frac{n}{2}} e^{-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\propto \lambda^{\frac{1}{2}}\lambda^{\alpha_0+\frac{n}{2}-1} e^{-\beta_0\lambda} e^{-\left(\frac{\lambda}{2}\right)\left[\kappa_0(\mu-\mu_0)^2+\sum_i(x_i-\mu)^2\right]}$$

Since $\sum_{i=1}^{n}(x_i-\mu)^2 = n(\mu-\bar{x})^2 + \sum_{i=1}^{n}(x_i-\bar{x})^2$

and $K_0(\mu-\mu_0)^2 + n(\mu-\bar{x})^2 = (K_0+n)(\mu-\mu_n)^2 + \frac{K_0 n(\bar{x}-\mu_0)^2}{K_0+n}$

where $\mu_n = \frac{K_0\mu_0 + n\bar{x}}{K_0+n}$

$$\Rightarrow K_0(\mu-\mu_0)^2 + \sum_i (x_i-\mu)^2 = K_0(\mu-\mu_0)^2 + n(\mu-\bar{x})^2 + \sum_i (x_i-\mu)^2$$

$$= (K_0+n)(\mu-\mu_n)^2 + \frac{K_0 n(\bar{x}-\mu_0)^2}{K_0+n} + \sum_i (x_i-\bar{x})^2$$

Therefore,

$$p(\mu,\lambda|D) \propto \lambda^{\frac{1}{2}} e^{-\left(\frac{\lambda}{2}\right)(K_0+n)(\mu-\mu_n)^2} \times \lambda^{\alpha_0+\frac{n}{2}-1} e^{-\beta_0\lambda} e^{-\left(\frac{\lambda}{2}\right)\sum_{i=1}^{N}(x_i-\bar{x})^2} e^{-\left(\frac{\lambda}{2}\right)\frac{K_0 n(\bar{x}-\mu_0)^2}{K_0+n}}$$

$$\propto \mathcal{N}\left(\mu \middle| \mu_n, ((K+n)\lambda)^{-1}\right) \times \text{Gam}\left(\lambda \middle| \alpha_0+\frac{n}{2}, \beta_n\right)$$

where $\beta_n = \beta_0 + \frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2 + \frac{K_0 n(\bar{x}-\mu_0)^2}{2(K_0+n)}$

In summary,

$$p(\mu,\lambda|D) = NG(\mu,\lambda|\mu_n, K_n, \alpha_n, \beta_n)$$

where $\mu_n = \frac{K_0\mu_0+n\bar{x}}{K_0+n}$ $\qquad \alpha_n = \alpha_0 + \frac{n}{2}$

$K_n = K_0+n$ $\qquad \beta_n = \beta_0 + \frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{x})^2 + \frac{K_0 n(\bar{x}-\mu_0)^2}{2(K_0+n)}$

plug in the values given:

$\mu_0 = 0, \; K_n = a^{-1}, \; \alpha_n = b, \; \beta_n = c$

$$\Rightarrow p(\mu,\lambda|D) = NG\left(\mu,\lambda \middle| \frac{n\bar{x}}{n+\frac{1}{a}}, \; n+\frac{1}{a}, \; \frac{n}{2}+b, \; c+\frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{x})^2 + \frac{n(\bar{x}-\mu_0)^2}{2a\left(n+\frac{1}{a}\right)}\right)$$

Problem 3. (b)

posterior : $p(\mu,\lambda \mid D) = NG\left(\mu,\lambda \,\middle|\, \frac{n\bar{x}}{n+\frac{1}{a}},\ n+\frac{1}{a},\ \frac{n}{2}+b,\ c+\frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{x})^2+\frac{n(\bar{x}-\mu_0)^2}{2a(n+\frac{1}{a})}\right)$

General form of predictive distribution:

$$p(x^* \mid D) = \int_0^{\infty}\int_{-\infty}^{\infty} p(x^* \mid \mu,\lambda)\, p(\mu,\lambda \mid D)\, d\mu\, d\lambda$$

$$= \int_0^{\infty}\int_{-\infty}^{\infty} \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(x^*-\mu)^2}{2}\right\} \cdot \frac{1}{Z_{NG}(\mu_n,K_n,\alpha_n,\beta_n)} \lambda^{\frac{1}{2}} \exp\left\{-\frac{K_n\lambda}{2}(\mu-\mu_n)^2\right\} \lambda^{\alpha_n-1} e^{-\lambda\beta_n}\, d\mu\, d\lambda$$

$$= \frac{1}{Z_{NG}\cdot\sqrt{2\pi}} \int_0^{\infty}\int_{-\infty}^{\infty} \lambda^{\alpha_n} \exp\left\{-\frac{\lambda(x^*-\mu)^2}{2} - \frac{K_n\lambda(\mu-\mu_n)^2}{2} - \lambda\beta_n\right\} d\mu\, d\lambda$$

$$= \frac{1}{Z_{NG}\cdot\sqrt{2\pi}} \int_0^{\infty}\int_{-\infty}^{\infty} \lambda^{\alpha_n} \exp\left\{-\frac{\lambda}{2}\left(x^{*2} - 2\mu x^* + \mu^2 + K_n\mu^2 - 2K_n\mu\cdot\mu_n + K_n\cdot\mu_n^2 + 2\beta_n\right\} d\mu\, d\lambda\right.$$

$$= \frac{1}{Z_{NG}\cdot\sqrt{2\pi}} \int_0^{\infty}\int_{-\infty}^{\infty} \exp\left\{-\frac{\lambda}{2}\left[\mu^2(1+K_n) - 2\mu(K_n\mu_n+x^*) + \frac{(K_n\mu_n+x^*)^2}{1+K_n}\right]\right. \cdot \lambda^{\alpha_n} \cdot \exp\left\{-\frac{\lambda}{2}\left(x^{*2} + K_n\cdot\mu_n^2 - \frac{(K_n\mu_n+x^*)^2}{1+K_n} + 2\beta_n\right)\right\} d\mu\, d\lambda$$

$$= \frac{1}{Z_{NG}\cdot\sqrt{2\pi}} \int_0^{\infty} \lambda^{\alpha_n} \exp\left\{-\frac{\lambda}{2}\left(x^{*2} + K_n\mu_n^2 - \frac{(K_n\mu_n+x^*)^2}{1+K_n} + 2\beta_n\right)\right\} \int_{-\infty}^{\infty} \exp\left\{-\left(\mu\sqrt{\frac{\lambda(1+K_n)}{2}} - \sqrt{\frac{\lambda}{2}}\cdot\left(\frac{K_n\mu_n+x^*}{\sqrt{1+K_n}}\right)\right)^2\right\} d\mu\, d\lambda$$

Assume $t = \left(\mu\sqrt{\frac{\lambda(1+K_n)}{2}} - \sqrt{\frac{\lambda}{2}}\cdot\left(\frac{K_n\mu_n+x^*}{\sqrt{1+K_n}}\right)\right)$ and $\int_{-\infty}^{\infty} \exp(-t^2)\,dt = \sqrt{\pi}$

$\Rightarrow dt = \sqrt{\frac{\lambda(1+K_n)}{2}}\, d\mu$ plug into the equation

$$\Rightarrow \frac{1}{Z_{NG}\cdot\sqrt{2}} \int_0^{\infty} \lambda^{\alpha_n} \cdot \sqrt{\frac{2}{\lambda(1+K_n)}} \cdot \exp\left\{-\frac{\lambda}{2}\left(x^{*2} + K_n\mu_n^2 - \frac{(K_n\mu_n+x^*)^2}{1+K_n} + 2\beta_n\right)\right\} d\lambda$$

let $S = \dfrac{x^{*2} + K_n\mu_n^2 - \frac{(K_n\mu_n+x^*)^2}{1+K_n} + 2\beta_n}{2}$

$r-1 = \alpha_n - \frac{1}{2}$

$$= \frac{1}{Z_{NG}\cdot\sqrt{2}} \cdot \sqrt{\frac{2}{1+K_n}} \cdot \int_0^{\infty} \lambda^{r-1} \cdot e^{-S\lambda}\, d\lambda$$

$$= \frac{1}{Z_{NG}\cdot\sqrt{2}} \sqrt{\frac{2}{1+K_n}} \cdot \frac{\Gamma(r)}{S^r} \underbrace{\int_0^{\infty} \frac{S^r}{\Gamma(r)} \lambda^{r-1} \cdot e^{-S\lambda}\, d\lambda}_{=1} = \underline{\underline{\frac{\Gamma(r)}{Z_{NG}\cdot S^r} \cdot \sqrt{\frac{1}{1+K_n}}}}$$

$$2S = x^{*2} + K_n\mu_n^2 - \frac{(K_n\mu_n+x^*)^2}{1+K_n} + 2\beta_n = \frac{K_n}{1+K_n}x^{*2} + \frac{K_n^2\mu_n^2 + K_n\mu_n^2 - K_n^2\mu_n^2 - 2x^*K_n\mu_n}{1+K_n} + 2\beta_n$$

$$= \frac{K_n}{1+K_n}x^{*2} - \frac{2x^*K_n\mu_n}{1+K_n} + \frac{K_n}{1+K_n}\mu_n^2 + 2\beta_n = \left(\frac{K_n}{1+K_n}\right)(x^*-\mu_n)^2 + 2\beta_n$$

$$Z_{NG}(\mu_n, K_n, \alpha_n, \beta_n) = \frac{\Gamma(\alpha_n)}{\beta_n^{\alpha_n}}\left(\frac{2\pi}{K_n}\right)^{\frac{1}{2}}$$

Combine all together:

$$\frac{\Gamma(r)}{Z_{NG} S^{r}} \cdot \sqrt{\frac{1}{1+K_n}} = \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \left(\frac{K_n}{2\pi}\right)^{\frac{1}{2}} \cdot \left[\beta_n + \frac{1}{2}\left(\frac{K_n}{K_n+1}\right)(x^* - \mu_n)^2 \right]^{-(\alpha_n+\frac{1}{2})} \cdot \Gamma(\alpha_n+\tfrac{1}{2}) \cdot \sqrt{\frac{1}{1+K_n}}$$

$$= \frac{\Gamma(\alpha_n+\frac{1}{2})}{\Gamma(\alpha_n)} \sqrt{\frac{1}{2\pi}\cdot\left(\frac{K_n}{1+K_n}\right)} \cdot \beta_n^{\alpha_n} \cdot \beta_n^{-(\alpha_n+\frac{1}{2})} \cdot \left[\frac{(x^*-\mu_n)^2}{\frac{2\beta_n(1+K_n)}{K_n}} + 1 \right]^{-(\alpha_n+\frac{1}{2})}$$

$$= \frac{\Gamma(\alpha_n+\frac{1}{2})}{\Gamma(\alpha_n)} \cdot \sqrt{\frac{1}{\pi}\cdot\frac{K_n}{2\beta_n(1+K_n)}} \cdot \left[1 + \frac{\alpha_n K_n (x^*-\mu_n)^2}{2\alpha_n\beta_n(1+K_n)} \right]^{-(\alpha_n+\frac{1}{2})}$$

$$= \frac{\Gamma(\alpha_n+\frac{1}{2})}{\Gamma(\alpha_n)} \cdot \sqrt{\frac{1}{\pi}} \cdot \sqrt{\frac{\alpha_n K_n}{2\cdot\alpha_n\beta_n(1+K_n)}} \left[1 + \frac{\alpha_n K_n (x^*-\mu_n)^2}{2\alpha_n\beta_n(1+K_n)} \right]^{-(\alpha_n+\frac{1}{2})}$$

Let $\Lambda = \dfrac{\alpha_n K_n}{\beta_n(K_n+1)} \Rightarrow p(x^*|D) = \dfrac{\Gamma(\alpha_n+\frac{1}{2})}{\Gamma(\alpha_n)} \cdot \pi^{-\frac{1}{2}} \cdot \left(\dfrac{\Lambda}{2\alpha_n}\right)^{\frac{1}{2}} \cdot \left(1 + \dfrac{\Lambda (x^*-\mu_n)^2}{2\alpha_n}\right)^{-(\alpha_n+\frac{1}{2})}$

This is Student's T-distribution, which center at μ_n with precision $\Lambda = \dfrac{\alpha_n K_n}{\beta_n(K_n+1)}$

Problem 4.

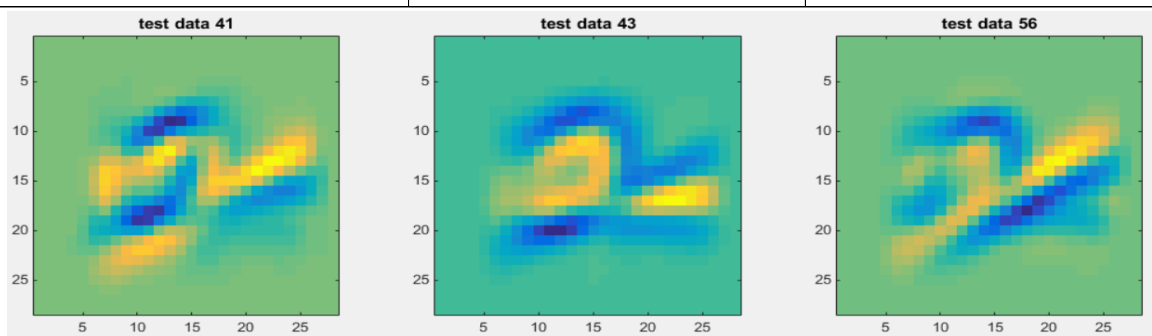
Details of implementing naïve Bayes classifier can be found in the codes.

Accuracy: 1857/1991 = 93.27%

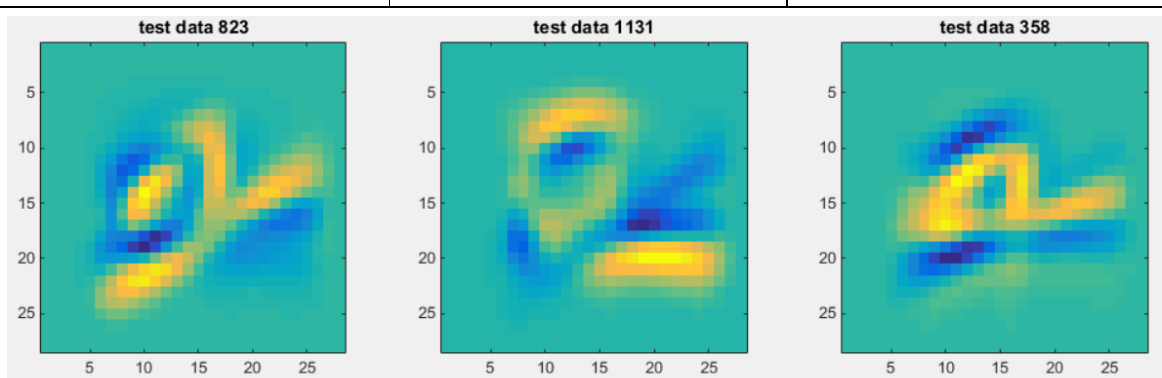
Confusion matrix:

	Classified as 0 (4)	Classified as 1 (9)
Label 0 (4)	930	52
Label 1 (9)	82	927

Misclassified:

Test data index	41	43	56
Images			
Predictive prob.	Pos: 1.07e-11/ Neg: 3.86e-12	Pos: 2.64e-10/ Neg: 2.44e-10	Pos: 3.26e-11/ Neg: 1.22e-11

Ambiguous:

Test data index	823	1131	358
Images			
Predictive prob.	Pos: 1.26e-11 / Neg: 1.26e-11	Pos: 2.33e-14 / Neg: 2.32e-14	Pos: 1.54e-11 / Neg: 1.56e-11
Ratio	0.4996	0.5020	0.4963