Bayesian Mod Machine Learning Assignment 1.

Haoyang Chen hc2812

1. Suppose D_i = the prize is behind ith door, $i=1,2,3$

S_i = She picks door i, $i=1,2,3$

H_i = Host opens door i, $i=1,2,3$

$$P(D_1|S_2,H_3) = \frac{P(H_3|D_1,S_2)P(D_1|S_2)P(S_2)}{P(S_2,H_3)} = \frac{P(H_3|D_1,S_2)P(D_1|S_2)P(S_2)}{\sum_{i=1}^{3} P(H_3|D_i,S_2)P(D_i|S_2)P(S_2)}$$

$$= \frac{P(H_3|D_1,S_2)}{\sum_{i=1}^{3} P(H_3|D_i,S_2)}$$

Since $P(H_3|D_1,S_2) = 1$, $P(H_3|D_2,S_2) = \frac{1}{2}$, $P(H_3|D_3,S_2) = 0$

$\therefore P(D_1|S_2,H_3) = \frac{1}{1+\frac{1}{2}+0} = \frac{2}{3}$

Similarly, $P(D_1|S_3,H_2) = P(D_2|S_1,H_3) = P(D_2|S_3,H_1) = P(D_3|S_1,H_2)$

$= P(D_3|S_2,H_1) = \frac{2}{3}$.

Conclusion, if she switch the door, $\frac{2}{3}$ chance could win the prize, otherwise only $\frac{1}{3}$ chance to win. Thus, $\boxed{\text{switch doors}}$

2. $X_i \sim$ multinomial(π), $i=1,\cdots,N$

$$P(\underline{X}|\pi) = \prod_{i=1}^{N} P(X_i|\pi) \propto \prod_{i=1}^{N}\prod_{j=1}^{k} \pi_j^{I(X_i=j)} = \prod_{j=1}^{k} \pi_j^{\sum_{i=1}^{N} I(X_i=j)}$$

Let $y_j = \sum_{i=1}^{N} I(X_i=j)$, the number of observations for category j, $j=1,\cdots,K$

Thus $P(\underline{X}|\pi) \propto \prod_{j=1}^{k} \pi_j^{y_j}$, the conjugate prior is Dirichlet Distribution

$$Dir(\pi|\alpha_1,\cdots,\alpha_K) = \frac{\Gamma(\alpha_1+\cdots+\alpha_K)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)} \prod_{j=1}^{k} \pi_j^{\alpha_j-1} \propto \prod_{j=1}^{k} \pi_j^{\alpha_j-1}$$

Thus the posterior $P(\underline{\pi}|\underline{x},\underline{\alpha}) \propto P(\underline{x}|\underline{\pi}) P(\underline{\pi}|\underline{\alpha}) \propto \prod_{j=1}^{K} \pi_j^{y_j} \cdot \prod_{j=1}^{K} \pi_j^{\alpha_j-1} = \prod_{j=1}^{K} \pi_j^{y_j+\alpha_j-1}$

Thus the posterior distribution is $Dir(\underline{\pi}|y_1+\alpha_1-1, \cdots, y_K+\alpha_K-1)$ Where $y_j = \sum_{i=1}^{N} I(x_i = j),\ j=1,\cdots,k$

The posterior distribution is a $\boxed{\text{Dirichlet Distribution}}$

The most important feature about the parameters is the parameters are the sum of prior parameters and counts of the observations.

3. a) $x_i \sim N(\mu, \lambda^{-1})$, $P(\underline{x}|\mu,\lambda) = \prod_{i=1}^{N} P(x_i|\mu,\lambda) \propto \prod_{i=1}^{N} \lambda^{\frac{1}{2}} \exp\{-\frac{(x_i-\mu)^2}{2}\lambda\} = \lambda^{\frac{N}{2}} \exp\{-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\lambda\}$

$\mu|\lambda \sim N(0, a\lambda^{-1})$, $P(\mu|\lambda) \propto \lambda^{\frac{1}{2}} \exp\{-\frac{\mu^2}{2a}\}$

$\lambda \sim Gamma(b,c)$, $P(\lambda) \propto \lambda^{b-1} \exp\{-c\lambda\}$

$$\therefore p(\mu,\lambda|\underline{x}) \propto P(\underline{x}|\mu,\lambda)\, p(\mu|\lambda)\, p(\lambda)$$

$$\propto \lambda^{\frac{N+1}{2}} \exp\{-\frac{\lambda(N+\frac{1}{a})}{2}\mu^2 + (\lambda\sum_{i=1}^{N}x_i)\cdot\mu - \frac{\lambda\sum_{i=1}^{N}x_i^2}{2}\}\ \lambda^{b-1}\exp\{-c\lambda\}$$

$$\propto \lambda^{\frac{N+1}{2}} \exp\{-\frac{\lambda(N+\frac{1}{a})}{2}(\mu - \frac{\sum_{i=1}^{N}x_i}{N+\frac{1}{a}})^2 - \lambda(\frac{\sum_{i=1}^{N}x_i^2}{2} - \frac{(\sum_{i=1}^{N}x_i)^2}{2(N+\frac{1}{a})})\}\ \lambda^{b-1}\exp\{-c\lambda\}$$

$$\propto \lambda^{\frac{1}{2}} \exp\{-\frac{\lambda(N+\frac{1}{a})}{2}(\mu - \frac{\sum_{i=1}^{N}x_i}{N+\frac{1}{a}})^2\}\ \lambda^{b+\frac{N}{2}-1}\exp\{-(c+\frac{\sum_{i=1}^{N}x_i^2}{2} - \frac{(\sum_{i=1}^{N}x_i)^2}{2(N+\frac{1}{a})})\lambda\}$$

$$\therefore P(\mu,\lambda|\underline{x}) = Normal(\mu|\frac{\sum_{i=1}^{N}x_i}{N+\frac{1}{a}}, \lambda(N+\frac{1}{a})) * Gamma(\lambda|b+\frac{N}{2}, c+\frac{\sum_{i=1}^{N}x_i^2}{2} - \frac{(\sum_{i=1}^{N}x_i)^2}{2(N+\frac{1}{a})})$$

Let $\mu^* = \frac{\sum_{i=1}^{N}x_i}{N+\frac{1}{a}}$, $\lambda^* = \lambda(N+\frac{1}{a})$, $\boxed{\mu|\lambda,\underline{x} \sim Normal(\mu^*, \lambda^{*-1})}$

Let $\alpha = b+\frac{N}{2}$, $\beta = c+\frac{\sum_{i=1}^{N}x_i^2}{2} - \frac{(\sum_{i=1}^{N}x_i)^2}{2(N+\frac{1}{a})} = c+\frac{1}{2}[(N-1)s^2 + \frac{N\bar{x}^2}{1+N}]$ where $s^2 = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N-1}$

$\boxed{\lambda|\underline{x} \sim Gamma(\alpha,\beta)}$

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

$$= \int_0^{\infty} p(\lambda|\underline{x}) \int_{-\infty}^{+\infty} p(x^*|\mu,\lambda)\, p(\mu|\lambda,\underline{x})\, d\mu\, d\lambda$$

$$= \int_0^{\infty} p(\lambda|\underline{x}) \int_{-\infty}^{+\infty} \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{\lambda(x^*-\mu)^2}{2}\right) \left(\frac{x^*}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{x^*(\mu-\mu^*)^2}{2}\right) d\mu\, d\lambda$$

$$= \int_0^{\infty} p(\lambda|\underline{x}) \int_{-\infty}^{+\infty} \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{\lambda(x^*-\mu)^2}{2}\right) \left(\frac{\lambda(N+\frac{1}{a})}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(N+\frac{1}{a})(\mu-\mu^*)^2}{2}\right\} d\mu\, d\lambda$$

$$= \int_0^{\infty} p(\lambda|\underline{x}) \int_{-\infty}^{+\infty} \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} \left(\frac{\lambda(N+\frac{1}{a})}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(x^*-\mu)^2}{2} - \frac{\lambda(N+\frac{1}{a})(\mu-\mu^*)^2}{2}\right\} d\mu\, d\lambda$$

Denote $N+\frac{1}{a} = \frac{1}{N^*}$. $\int_{-\infty}^{+\infty} \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} \left(\frac{\lambda(N+\frac{1}{a})}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(x^*-\mu)^2}{2} - \frac{\lambda(N+\frac{1}{a})(\mu-\mu^*)^2}{2}\right\} d\mu$

$$= \int_{-\infty}^{+\infty} \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} \left(\frac{\lambda}{2\pi N_0}\right)^{\frac{1}{2}} \left(\frac{\lambda(1+\frac{1}{N^*})}{2\pi}\right)^{-\frac{1}{2}} \left(\frac{\lambda(1+\frac{1}{N^*})}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda}{2}\left(1+\frac{1}{N^*}\right)\left(\mu - \frac{\lambda x^* + \frac{\lambda\mu^*}{N^*}}{\lambda(1+\frac{1}{N^*})}\right)^2\right\}$$

$$* \exp\left\{-\frac{\lambda(x^*)^2}{2} - \frac{\lambda\mu^{*2}}{2N^*} + \frac{(\lambda x^* + \frac{\lambda\mu^*}{N^*})^2}{2\lambda(1+\frac{1}{N^*})}\right\} d\mu$$

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

Since $\int_{-\infty}^{+\infty} \left(\frac{\lambda(1+\frac{1}{N^*})}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(1+\frac{1}{N^*})}{2}\left(\mu - \frac{\lambda x^* + \frac{\lambda\mu^*}{N^*}}{\lambda(1+\frac{1}{N^*})}\right)^2\right\} d\mu = \int_{-\infty}^{+\infty} \mathrm{Normal}\left(\mu \,\middle|\, \frac{\lambda x^* + \frac{\lambda\mu^*}{N^*}}{\lambda(1+\frac{1}{N^*})}, \left(\lambda(1+\frac{1}{N^*})\right)^{-1}\right) d\mu = 1$

$$\therefore p(x^*|\underline{x}) = \int_0^{\infty} p(\lambda|\underline{x}) \cdot \left(\frac{\lambda}{2\pi(N^*+1)}\right)^{\frac{1}{2}} \exp\left(-\frac{\lambda(x^*-\mu^*)^2}{2(N^*+1)}\right) d\lambda$$

$$= \int_0^{\infty} \frac{\beta^{\alpha}\lambda^{\alpha-1}\exp(-\beta\lambda)}{\Gamma(\alpha)} \left(\frac{\lambda}{2\pi(N^*+1)}\right)^{\frac{1}{2}} \exp\left(-\frac{\lambda(x^*-\mu^*)^2}{2(N^*+1)}\right) d\lambda$$

$$= \int_0^{\infty} \frac{\beta^{\alpha}\lambda^{\alpha+\frac{1}{2}-1}\exp\left(-\lambda\left(\beta+\frac{(x^*-\mu^*)^2}{2(N^*+1)}\right)\right)}{\Gamma(\alpha)} \cdot \left(\frac{1}{2\pi(N^*+1)}\right)^{\frac{1}{2}} d\lambda$$

$$= \frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)} \cdot \left(\frac{1}{2\pi(N^*+1)}\right)^{\frac{1}{2}} \cdot \frac{\beta^{\alpha}}{\left(\beta+\frac{(x^*-\mu^*)^2}{2(N^*+1)}\right)^{\alpha+\frac{1}{2}}} \int_0^{\infty} \frac{\left(\beta+\frac{(x^*-\mu^*)^2}{2(N^*+1)}\right)^{\alpha+\frac{1}{2}} \cdot \lambda^{\alpha+\frac{1}{2}-1} \exp\left(-\left(\beta+\frac{(x^*-\mu^*)^2}{2(N^*+1)}\right)\lambda\right)}{\Gamma(\alpha+\frac{1}{2})} d\lambda$$

Since $\int_0^{\infty} \left(\frac{\left(\beta+\frac{(x^*-\mu^*)^2}{2(N^*+1)}\right)^{\alpha+\frac{1}{2}} \cdot \lambda^{\alpha+\frac{1}{2}-1} \exp\left(-\left(\beta+\frac{(x^*-\mu^*)^2}{2(N^*+1)}\right)\lambda\right)}{\Gamma(\alpha+\frac{1}{2})}\right) d\lambda = \int_0^{\infty} \mathrm{Gamma}\left(\lambda \,\middle|\, \alpha+\frac{1}{2}, \beta+\frac{(x^*-\mu^*)}{2(N^*+1)}\right) d\lambda = 1$

$$\therefore p(x^*|\underline{x}) = \frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)} \cdot \left(\frac{1}{2\pi(N^*+1)}\right)^{\frac{1}{2}} \cdot \frac{\beta^{\alpha}}{\left(\beta+\frac{(x^*-\mu^*)^2}{2(N^*+1)}\right)^{\alpha+\frac{1}{2}}}$$

$$= \frac{\Gamma(\frac{2\alpha+1}{2})}{\Gamma(\frac{2\alpha}{2})} \cdot \left(\frac{1}{\pi\cdot 2\alpha}\right)^{\frac{1}{2}} \left(\frac{1}{\beta(N^*+1)/\alpha}\right)^{\frac{1}{2}} \cdot \left[1+\frac{1}{2\alpha}\cdot\frac{\alpha}{\beta(N^*+1)}\cdot(x^*-\mu^*)^2\right]^{-\frac{2\alpha+1}{2}}$$

$\sim t_{2\alpha}\left(x^* \,\middle|\, \mu^*, \frac{\beta(N^*+1)}{\alpha}\right)$: a student's t-distribution with $df=2\alpha$, $loc=\mu^*$, $scale=\frac{\beta(N^*+1)}{\alpha}$

where $\mu^* = \frac{\sum_{i=1}^{N} x_i}{N+\frac{1}{a}}$. $\alpha = b+\frac{N}{2}$, $\beta = c+\frac{1}{2}\left[(N-1)s^2+\frac{N\bar{x}^2}{1+N}\right]$, $s^2 = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N-1}$. $N^* = \frac{1}{N+\frac{1}{a}}$

4.

a). Shown in the code

b). The cross table is shown below:

	0	1
0	930	52
1	82	927

Accuracy = 0.9327

c). 3 misclassified images

Image Index: 223

True label: 0

Predicted label: 1

Logit: 0.639793214237

$P(y^* = 0 \mid x^*, X, y) = 0.3902$

$P(y^* = 1 \mid x^*, X, y) = 0.6098$

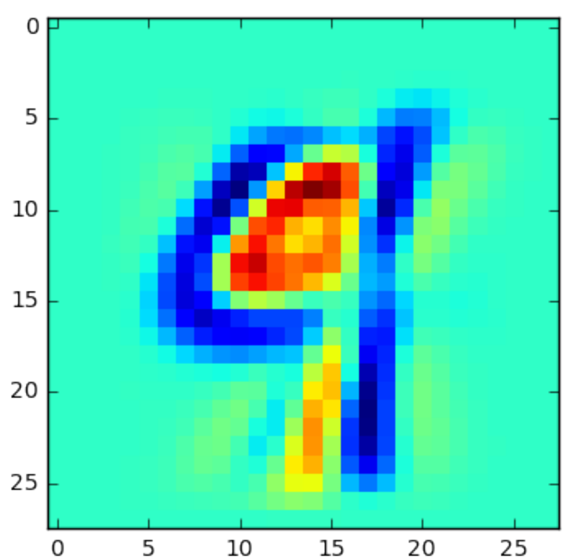


Image Index: 676
True label: 0
Predicted label: 1
Logit: 0.926222661075
$P(y^* = 0 \mid x^*, X, y) = 0.4808$
$P(y^* = 1 \mid x^*, X, y) = 0.5192$

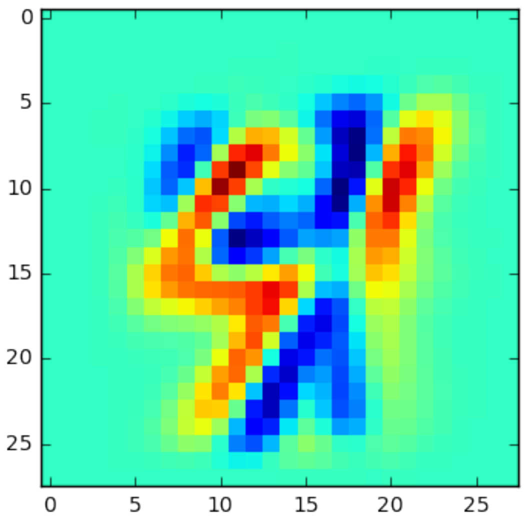
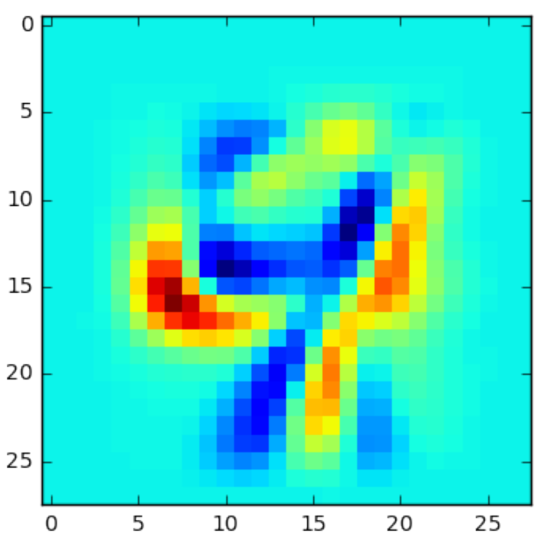


Image Index: 1842
True label: 1
Predicted label: 0
Logit: 1.79610067819
$P(y^* = 0 \mid x^*, X, y) = 0.6424$
$P(y^* = 1 \mid x^*, X, y) = 0.3576$



d). 3 most ambiguous predictions
Image Index: 822
True label: 0
Predicted label: 0
Logit: 1.00118731725
$P(y^* = 0 \mid x^*, X, y) = 0.5003$
$P(y^* = 1 \mid x^*, X, y) = 0.4997$

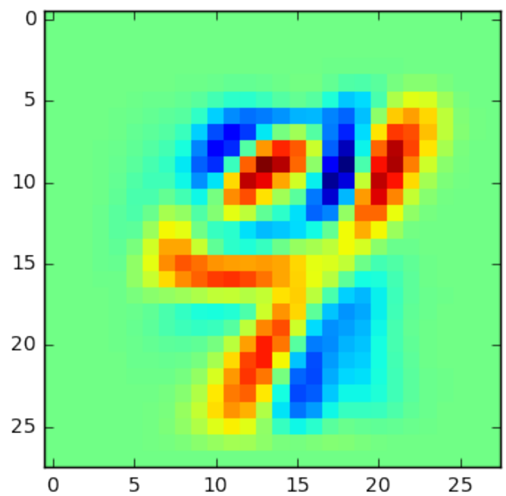


Image Index: 1130
True label: 1
Predicted label: 1
Logit: 0.990875806066
$P(y^* = 0 \mid x^*, X, y) = 0.4977$
$P(y^* = 1 \mid x^*, X, y) = 0.5023$

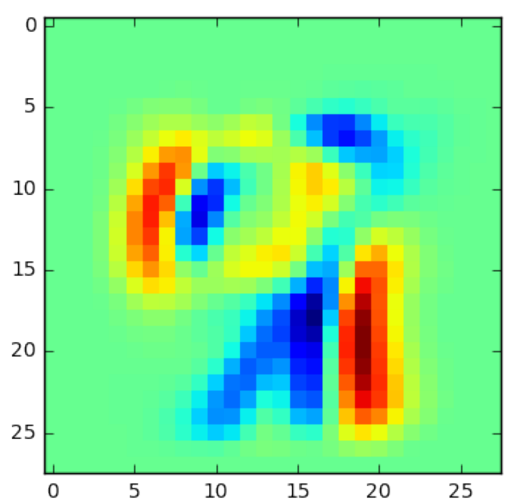


Image Index: 357
True label: 0
Predicted label: 0
Logit: 1.01456469643
$P(y^* = 0 \mid x^*, X, y) = 0.5036$
$P(y^* = 1 \mid x^*, X, y) = 0.4964$