Bayesian Machine learning HW1

Benjamin Zhu
bjz2107

1) 3 doors A B C behind 1 = A behind 2 = B behind 3 = C

$P(\text{prize}) = \frac{1}{3} = P(A) = P(B) = P(C)$ guy opening door 1 = X door 2 = Y door 3 = Z

After making initial choice of picking door 1 the remaining door to be opened

$$P(A|Z) = \frac{P(Z|A)P(A)}{P(Z)}$$

prior $P(Z|A) = \frac{1}{2}$ because there are two other doors to open

$P(Z|C) = 0$

$P(Z|B) = 1$

posterior = $\frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$

$$P(A|Z) = \frac{P(Z|A)\left(\frac{1}{3}\right)}{\left(\frac{1}{2}\right)}$$

$$= \frac{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)}{\left(\frac{1}{2}\right)} = \boxed{\frac{1}{3}}$$ probability behind 1, given he opens door 3

$$P(B|Z) = \frac{P(Z|B)P(B)}{P(Z)}$$

$P(Z|B) = 1$

$P(Z) = \frac{1}{2}$

$$= \frac{(1)\left(\frac{1}{3}\right)}{\left(\frac{1}{2}\right)} = \boxed{\frac{2}{3}}$$ prob behind 2, given opens door 3

switch doors! if he opens a door

2) $\pi = (\pi_1 \ldots \pi_K)$ $\pi_j \geq 0$ $\Sigma_j \pi_j = 1$ for $i = 1 \ldots N$

Multinomial distribution

$x_i \sim$ Multinomial (π) iid for $i = 1 \ldots N$

$$p(x_i \mid \pi) = \frac{\Gamma(n+1)}{\prod_{j=1}^{K} \Gamma(x_j+1)} \prod_{j=1}^{K} \pi_j^{x_j} \quad \text{for } i = 1 \ldots N$$

for an ordinary multinomial

Generalize for all x_i's

$$p(x_1 \ldots x_N \mid \pi) = \prod_{i=1}^{N} \Bigg[\underbrace{\frac{\Gamma(n+1)}{\prod_{j=1}^{K} \Gamma(x_j+1)}}_{\text{normalizing constant}} \prod_{j=1}^{K} \pi_j^{x_j} \Bigg]$$

now as we saw in class

$$\propto \prod_{i=1}^{N} \prod_{j=1}^{K} \pi_j^{x_{ij}}$$

Conjugate prior to multinomial is the Dirichlet distribution

$$p(x_i \mid \alpha_1 \ldots \alpha_K) = \underbrace{\frac{\Gamma \sum_{j=1}^{K} \alpha_j}{\Gamma(\alpha_1)\Gamma(\alpha_2) \ldots \Gamma(\alpha_K)}} \prod_{j=1}^{K} x_i^{\alpha_j - 1}$$

and we know the normalizing constant can drop out

$$\propto \prod_{j=1}^{K} x_i^{\alpha_j - 1}$$

$$p(\pi_j \mid \alpha_1 \ldots \alpha_K) \propto \prod_{j=1}^{K} \pi_j^{\alpha_j - 1}$$

for each individual π_j, as we are looking for the prior + posterior

2) continued)   posterior of $(\theta | x)$

$$p(\pi | x_{i \ldots N}, \alpha_{1 \ldots K}) = \frac{p(x_{i \ldots N} | \pi) p(\pi)}{\int p(x_{i \ldots N} | \pi) p(\pi) d\pi}$$

$$= \frac{\left[\prod_{i=1}^{N} \left(\frac{\Gamma(n+1)}{\prod_{j=1}^{K} \Gamma(x_j + 1)} \prod_{j=1}^{K} \pi_j^{x_{ij}} \right)\right] \left[\frac{\Gamma \sum_{j=1}^{K} \alpha_j}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_K)} \prod_{j=1}^{K} \pi_j^{\alpha_j - 1} \right]}{\int \text{something as above, } d\pi}$$

constants drop out as before

$$\propto \underbrace{\left(\prod_{i=1}^{N} \prod_{j=1}^{K} \pi_j^{x_{ij}} \right)}_{\prod_{j=1}^{K} \pi_j^{\sum_{i=1}^{N} x_{ij}}} \left(\prod_{j=1}^{K} \pi_j^{\alpha_j - 1} \right)$$

$\prod_{j=1}^{K} \pi_j^{\sum_{i=1}^{N} x_{ij}}$ where $\sum_{i=1}^{N} x_i$ is a hit on the item K

$$\propto \prod_{j=1}^{K} \pi_j^{\alpha_j + \sum_{i=1}^{N} x_{ij} - 1}$$ This is a Dirichlet distribution

$$Dir\left(\alpha_1 + \sum_{\substack{i=1 \\ i=j}}^{N} x_{ij}, \ldots \alpha_K + \sum_{\substack{i=1 \\ i=K}}^{N} x_{ij}\right)$$

This is the same distribution as the prior, though with some changes → just a Beta normalizing constant in front of the outside

3) $(x_1 \dots x_N)$ $x \in \mathbb{R}$ modelled as iid, $\text{Normal}(\mu, \lambda^{-1})$

$\mu | \lambda \sim \text{Normal}(0, a\lambda^{-1})$ and $\lambda \sim \text{Gamma}(b, c)$

Ok we know both of these are exponential, so they will integrate through

$$x_i \overset{iid}{\sim} \text{Normal}(\mu, \lambda^{-1})$$

$$\text{Normal}(x_i | \mu, \lambda^{-1}) = (2\pi\lambda^{-1})^{\frac{n}{2}} \exp\left(-\frac{1}{2\lambda^{-1}}(x_i - \mu)^2\right\}$$

$$p(x_1 \dots x_N | \mu, \lambda) = \prod_{i=1}^{N} p(x_i | \mu, \lambda)$$

$$p(\mu, \lambda | x_i) = \frac{p(x_i | \mu, \lambda)\, p(\mu, \lambda)}{p(x_i)} = \frac{p(x_i | \mu, \lambda)\, p(\mu, \lambda)}{\iint p(x_i | \mu, \lambda)\, p(\mu, \lambda)\, d\mu\, d\lambda}$$

$$p(\mu, \lambda) = p(\mu | \lambda)\, p(\lambda)$$

$$= \frac{p(x_i | \mu, \lambda)\, p(\mu | \lambda)\, p(\lambda)}{\iint p(x_i | \mu, \lambda)\, p(\mu | \lambda)\, p(\lambda)\, d\mu\, d\lambda}$$

HMM okay, so

$$p(x_{1 \dots N} | \mu, \lambda^{-1}) = \prod_{i=1}^{N} (2\pi\lambda^{-1})^{-\frac{n}{2}} \exp\left[-\frac{1}{2\lambda^{-1}}(x_i - \mu)^2\right]$$

$$p(\mu | \lambda) = (2\pi a\lambda^{-1})^{-\frac{1}{2}} \exp\left[\frac{-\lambda}{2a}(\mu)^2\right]$$

$$p(\lambda) = \frac{c^b}{\Gamma(b)} \lambda^{b-1} e^{-c\lambda}$$

just distributions

just so we can solve for something

$$p(\mu | \lambda, x_i) = \frac{p(\mu | \lambda)\, p(x_i | \mu | \lambda)}{\int p(x_i | \mu | \lambda)\, p(\mu | \lambda)} \propto p(\mu | \lambda)\, p(x_i | \mu | \lambda)$$

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

$$\propto \left(2\pi \frac{a}{\lambda}\right)^{-\frac{1}{2}} \left(2\pi \frac{1}{\lambda}\right)^{-\frac{n}{2}} e\left[\frac{-\lambda}{2a}\mu^2 - \frac{\lambda}{2} \sum_{i=1}^{n} (x_i - \mu)^2\right]$$

... close but one sec, will backtrack

cont ...

cont inued

3) continued

ok so wikipedia has the Normal Gamma to get μ, λ joint

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

$$= (2\pi a)^{-\frac{1}{2}} \frac{c^b}{\Gamma(b)} \lambda^{b-1+\frac{1}{2}} \exp\left[-c\lambda\right] \exp\left[-\frac{1}{2a\lambda^{-1}}(\mu)^2\right]$$

this is the normal gamma distribution for μ, λ

$x_i = N(\mu, \lambda^{-1})$

Posterior $p(x_{1:N} \mid \mu, \lambda)\, p(\mu, \lambda)$

$$\propto \prod_{i=1}^{N} (2\pi\lambda^{-1})^{-\frac{1}{2}} \exp\left[-\frac{1}{2\lambda^{-1}}(x_i - \mu)^2\right] \frac{c^b}{\Gamma(b)} (2\pi a)^{-\frac{1}{2}} \lambda^{b-\frac{1}{2}} \exp[-c\lambda] \cdot \exp\left[\frac{1}{2a\lambda}\mu^2\right]$$

$$\propto (2\pi\lambda^{-1})^{-n/2} \exp\left[-\frac{1}{2}\lambda \sum^{n} (x_i - \mu)^2\right] \left(\frac{c^b}{\Gamma(b)} (2\pi a)^{-\frac{1}{2}} \lambda^{b-\frac{1}{2}} \exp[-c\lambda] \exp\left[-\frac{1}{2a\lambda^{-1}}(\mu^2)\right]\right)$$

$$\propto \frac{c^b}{\Gamma(b)(2\pi)^{\frac{n+1}{2}} a^{-\frac{1}{2}}} \lambda^{(b+\frac{n}{2})-\frac{1}{2}} \exp[-c\lambda] \exp\left[-\frac{1}{2}\lambda\left(a(\mu)^2 + \sum^{n}(x_i - \mu)^2\right)\right]$$

so

we're getting there, need to complete the square $a(\mu)^2 + \sum (x_i - \mu)^2$

also, unclear if a is a^{-1} or not, but it doesn't actually matter...

$$\sum (x_i - \mu)^2 = \sum (x_i + \bar{x} - \bar{x} - \mu^2) = \sum (\bar{x} - \mu + x_i - \bar{x})^2$$

$$= \sum (\bar{x} - \mu)^2 + \sum^{n} (x_i - \bar{x})^2$$

this is just the variance,

$$= n(\bar{x} - \mu)^2 + n\sigma$$

$\sigma = \frac{1}{n} \sum (x_i - \bar{x})^2$

complete this square

$$a(\mu)^2 + n\sigma + n(\bar{x} - \mu)^2$$

$$= a\mu^2 + n\sigma + n\bar{x} - n2\bar{x}\mu + n\mu^2$$

continued

3) continued 2

$$a\mu^2 + n\sigma + n\bar{x}^2 - n2\mu\bar{x} - n\mu^2 = n\sigma + n\bar{x}^2 + a\mu^2 - n2\mu\bar{x} + n\mu^2$$

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

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

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

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

$$n\bar{x}^2(a+n)$$
$$an\bar{x}^2 + n^2\bar{x}^2 - n\bar{x}^2$$

Back to posterior

$$\propto \frac{c^b}{\Gamma(b)}(2\pi)^{-\left(\frac{n+1}{2}\right)} a^{-\frac{1}{2}} \lambda^{\left(b+\frac{n}{2}\right)-\frac{1}{2}} \exp\left[-c\lambda\right]\exp\left[-\frac{1}{2}\lambda\left((a+n)\left(\mu - \frac{n\bar{x}}{a+n}\right)^2 + n\sigma + n\bar{x}^2\left(\frac{2n+a}{n+a}\right)\right)\right]$$

$$\propto \frac{c^b}{\Gamma(b)}(2\pi)^{-\left(\frac{n+1}{2}\right)} a^{-\frac{1}{2}} \lambda^{\left(b+\frac{n}{2}\right)-\frac{1}{2}} \exp\left[-\frac{1}{2}\lambda(a+n)\left(\mu - \frac{n\bar{x}}{a+n}\right)^2\right]\exp\left[-\lambda\left[c + \frac{n\bar{x}^2}{2(a+n)}(2n+a) + \frac{n\sigma}{2}\right]\right]$$

so uhhh..

new parameters

$$\hat{a} = a+n \qquad \hat{c} = c + \frac{1}{2}\left(\frac{n\bar{x}^2(2n+a)}{a+n}\right) + n\sigma$$

$$\hat{b} = b + \frac{n}{2} \qquad \hat{\mu} = \left(\frac{n\bar{x}}{a+n}\right) \qquad \text{I think} \ldots$$

$$p(\mu,\lambda \mid a,b,c,X) = \cancel{\frac{c^b}{\Gamma(b)}(2\pi)^{-\frac{n+1}{2}} a^{-\frac{1}{2}}} \lambda^{\hat{b}-\frac{1}{2}} \exp\left[-\lambda\hat{c}\right]\exp\left[-\frac{1}{2}\lambda\hat{a}(\mu-\hat{\mu})^2\right]$$

$\iint$ of above $d\mu\, d\lambda$

$$\propto \lambda^{\hat{b}-\frac{1}{2}} \exp\left[-\lambda\hat{c}\right]\exp\left[-\frac{1}{2}\lambda\hat{a}(\mu-\hat{\mu})^2\right] \cdot \frac{\hat{c}^{\hat{b}}}{\Gamma(\hat{b})}(2\pi\hat{a})^{-\frac{1}{2}}$$

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

$$= \iint p(x^*|\mu,\lambda)\, p(\mu|a,\lambda,x_i\text{'s})\, p(\lambda|b,c,x_i\text{'s})\, d\mu\, d\lambda$$

actually... $p(x^*|x_1 \ldots x_N) = \iint p(x^*|\mu,\lambda)\, p(x|\mu,\lambda)\, p(\mu|\lambda)\, p(\lambda)\, d\mu\, d\lambda$

$$= \iint \frac{c^b}{\Gamma(b)} \lambda^{b-\frac{1}{2}} \exp[-c\lambda]\, \lambda^{\frac{n}{2}+\frac{1}{2}} (2\pi)^{-(\frac{n+1}{2})-\frac{1}{2}} a^{-\frac{1}{2}} \exp\left[-\frac{\lambda}{2}\sum(x_i-\mu)^2\right] \exp\left[\frac{-\lambda}{2a}\mu^2\right] \exp\left[(x^*-\mu)^2\frac{\lambda}{2}\right] d\mu\, d\lambda$$

same thing

ok this is absurdly long...

μ $d\mu$ first

going to do some handwaving

$$\int_\mu p(x^*|\mu,\lambda)\, p(\mu|x_i,\lambda)\, d\mu$$

$$= \int (2\pi\lambda^{-1})^{-\frac{1}{2}} \exp\left[-\frac{\lambda}{2}(x^*-\mu)^2\right] (2\pi a\lambda^{-1})^{-\frac{1}{2}} \exp\left(-\frac{1}{2a\lambda^{-1}}\mu^2\right] d\mu$$

$$\propto \exp\left[\frac{-\lambda}{2}\left((n\bar{x})^2 + x^{*2} - 2\mu n\bar{x} + n\mu^2 + \mu^2 - 2\mu x^* + \frac{\mu^2}{a}\right]\right] d\mu$$

$$= \int \exp\left[\frac{-\lambda}{2}\left((n\bar{x})^2 + x^{*2} + \left(\frac{2n+a}{a}\right)\mu^2 - 2\mu(n\bar{x}+x^*)\right)\right]$$

integrate

$$\propto \frac{(2\pi a)^{-\frac{1}{2}}}{(2n+a)^{-\frac{1}{2}}} \exp\left[\frac{-\lambda\left[\left(\frac{2n+a}{a}\right)n\bar{x})^2 + \left(\frac{2n+a}{a}\right)x^{*2} - (n\bar{x})^2 - x^{*2} - 2n\bar{x}x^*\right]}{2\left(\frac{2n+a}{a}\right)}\right]$$

$d\mu$ is out

$$p(x^*|x_1 \ldots x_N) \propto \int \lambda^{\frac{n-1}{2}+b} \exp\left[-\lambda\left[\frac{(na+1)x^{*2} - 2a\bar{x}x^*}{2\left(\frac{2n+a}{a}\right)} + \frac{(2n+1)(n\bar{x})^2}{2(2n+a)} + c\right] d\lambda\right.$$

$$\propto \int \lambda^{b+\frac{n-1}{2}} \exp\left[-\frac{\lambda}{2n+c}\left((na+1)x^{*2} - 2x^*(\bar{x}an + c + (n+1)n\bar{x}^2\right)\right)$$

$$\propto \int \lambda^{\frac{n-1}{2}+b} \exp\left[\frac{-\lambda}{(2n+c)}\left(x^{*2}\left(1+\frac{1}{na}\right) - 2x^*\bar{x} + c + na + (n\bar{x})^2\right)\right] an\, d\lambda$$

this is not dependent on λ.

3) b) continued

So ... continuing forward,

the expression looks like we're getting there.

$$\int \lambda^{\frac{n-1}{2}+b} \exp\left[\frac{-\lambda}{2(2n+a)}(1+an)\left(x^* - \frac{n\bar{x}}{a+n}\right)^2 + 1\right] d\lambda$$

This is almost the Γ

given

$$\int_{-\infty}^{\infty} x^{\alpha-1} \exp[-\beta x]\, dx = \frac{\Gamma(\alpha)}{\beta^{\alpha}}$$

$$p(x^*|x_1 \ldots x_n) = \frac{c^b}{\Gamma(b)}(2\pi)^{-\left(\frac{n+1}{2}\right)}(2n+a)^{-\frac{1}{2}}\,\Gamma\left(\frac{n+1}{2}+c\right)\left(\frac{(na+1)\left(x^* - \frac{n\bar{x}}{a+n}\right)}{\frac{(2n+a)}{a}}+1\right)^{b+\frac{1}{2}}$$

$$= \frac{c^b}{\Gamma(b)} 2\pi^{-\left(\frac{n+1}{2}\right)}(2n+a)^{-\frac{1}{2}}\,\Gamma\left(\frac{n+1}{2}+c\right)$$

ok so I messed up some constants somewhere, BUT, this is

$$\propto \frac{\Gamma\left(\frac{n+1}{2}+c\right)}{\Gamma(b)} \frac{\sqrt{a}\;c^b}{\sqrt{2n+a}} \left(\frac{(x^*-n)^2}{a+1}\frac{a}{2}+c\right)^{b+\frac{1}{2}}$$

which is a student t distribution

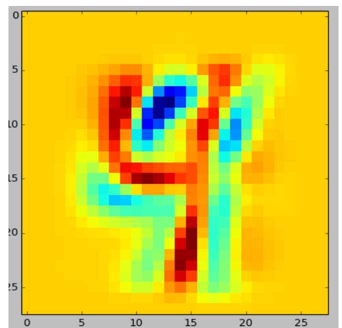
4)a) done, code attached

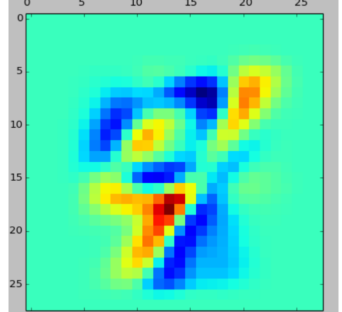
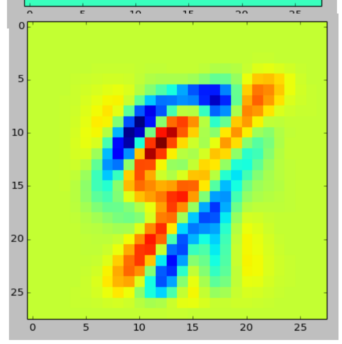
4)b)

```
Done
Predicted    0    1
Actual
0          939   43
1           92  917

[2 rows x 2 columns]
Accuracy is
Model accuracy is 0.932194876946
1856/1991
```

4)c)

```
These three are wrong
0                       -3.28127
1                      -3.192847
Predicted                      1
Actual                         0
Probability_y0         0.4931711
Probability_y1         0.5068289
Wrong                          1
Ambiguous            0.006828926
Name: 40, dtype: object
0                       -2.89678
1                      -2.879747
Predicted                      1
Actual                         0
Probability_y0         0.4985256
Probability_y1         0.5014744
Wrong                          1
Ambiguous            0.001474352
Name: 55, dtype: object
0                      -2.76888
1                     -2.516782
Predicted                     1
Actual                        0
Probability_y0        0.4761526
Probability_y1        0.5238474
Wrong                         1
Ambiguous            0.02384737

Name: 64, dtype: object
```



4)d)

```
Three most ambiguous
0                    -3.535112
1                    -3.537115
Predicted                    0
Actual                       1
Probability_y0       0.5001416
Probability_y1       0.4998584
Wrong                        1
Ambiguous         0.0001416222
Name: 1988, dtype: object
0                    -1.965257
1                    -1.963564
Predicted                    1
Actual                       0
Probability_y0       0.4997846
Probability_y1       0.5002154
Wrong                        1
Ambiguous         0.0002153807
Name: 195, dtype: object
0                    -2.981896
1                    -2.984986
Predicted                    0
Actual                       0
Probability_y0        0.500259
Probability_y1        0.499741
Wrong                        0
Ambiguous         0.0002589667
Name: 81, dtype: object
```