Changxin Miao
11853018
changxin.miao@student.uva.nl

Homework Assignment 1
Machine Learning 2, 17/18

2018-04-11

1 Problem 1

1.1 1

The μ_y of the Gaussian distribution will be the expectation of the distribution

$$\mu_y = E_y = E(x+y) \tag{1}$$
$$E_y = E_x + E_z \tag{2}$$
$$E_y = \mu_x + \mu_z \tag{3}$$

$$\begin{aligned}
Cov[y,y] &= E[yy^T] - E[y]E[y^T] && (4)\\
&= E[(x+z)(x+z^T)] - E[(x+z)]E[(x+z)^T] && (5)\\
&= E[xx^T + zx^T + xz^T + ZZ^t] - (\mu_x+\mu_z)(\mu_x+\mu_z)^T && (6)\\
&= E[xx^T] + E[zx^T] + E[xz^T] + E[zz^T] - (\mu_x+\mu_z)(\mu_x+\mu_z)^T && (7)
\end{aligned}$$

With the independence assumption for μ_x and μ_z, we can derive that

$$\begin{aligned}
Cov[y,y] &= E[xx^T] + E[z]E[x^T] + E[x]E[z^T] + E[zz^T] - (\mu_x+\mu_z)(\mu_x+\mu_z)^T && (8)\\
&= E[xx^T] + 2\mu_x\mu_z + E[zz^T] - (\mu_x\mu_x^T + 2\mu_z\mu_x^T + \mu_z\mu_z^T) && (9)\\
&= E[xx^T] - \mu_x\mu_x^T + E[zz^T] - \mu_z\mu_z^T && (10)\\
&= E[xx^T] - E[x]E[x^T] + E[zz^T] - E[z]E[z^T] \qquad = Cov[x,x] + Cov[z,z] && (11)\\
&= \Sigma_x + \Sigma_z && (12)
\end{aligned}$$

2 2

2.1 1

We have a normal distribution for x and the Σ is known. Now we need to derive the log-likelihood based on it:

$$\mathcal{N}(x \mid \mu, \Sigma) = \prod_{n=1}^{N} \mathcal{N}(x_n \mid \mu, \Sigma) \tag{13}$$

2.2 2

$$\begin{aligned}
p(\mu \mid x, \Sigma, \mu_0, \Sigma_0) &= \frac{p(x \mid \mu, \Sigma)p(\mu \mid \mu_0, \Sigma_0)}{p(x)} && (14)\\
&\propto \prod_{n=1}^{N} \mathcal{N}(x_n \mid \mu, \Sigma)\mathcal{N}(x_n \mid \mu_0, \Sigma_0) && (15)
\end{aligned}$$

$$p(\mu \mid x, \Sigma, \mu_0, \Sigma_0) \propto \prod_{n=1}^{N} \mathcal{N}(x_n \mid \mu, \Sigma)\mathcal{N}(x_n \mid \mu_0, \Sigma_0) \propto \prod_{n=1}^{N} (\frac{1}{(2\pi)^{\frac{1}{2}} \mid \Sigma \mid^{\frac{1}{2}}} exp(-\frac{1}{2}(x_n - \mu)^T \Sigma^{-1} \tag{16}$$

$$\propto C(exp \sum_{n=1}^{N} (-\frac{1}{2}(x_n^T \Sigma^{-1} x_n - 2\mu\Sigma^{-1} x_n + \tag{17}$$

$$\sum_{n=1}^{N} (-\frac{1}{2}(x_n^T \Sigma^{-1} x_n - 2\mu\Sigma^{-1} x_n + \mu^T \Sigma^{-1} \mu) - \frac{1}{2}(\mu^T \Sigma_0^{-1} \mu^T - 2\mu_0^T \Sigma_0^{-1} \mu + \mu_0^T \Sigma_0^{-1} \mu_0 \tag{18}$$

$$\propto \mu^T (\Sigma_0^{-1} - \Sigma^{-1})\mu - 2(\Sigma \sum_{n=1}^{N} xn + \mu_0^T \Sigma_0 \tag{19}$$

As the posterior distribution is a Gaussian distribution $\mathcal{N}(\mu \mid \mu_N, \Sigma_N)$

$$\mathcal{N}(\mu \mid \mu_N, \Sigma_N) \propto \mu^T \Sigma_N^{-1} \mu_N - 2\mu_N \Sigma_N^{-1} \mu + \mu_N \mu_N^T \tag{20}$$

Compare the above two equations ,we could conclude that

$$\Sigma_N^{-1} = \Sigma_0^{-1} - \Sigma^{-1} \tag{21}$$

$$\mu_N \Sigma_N^{-1} \mu = (\sum_{n=1}^{N} x_n \Sigma^{-1} - \mu_0 \Sigma_0^{-1})\mu \tag{22}$$

$$\mu_N \Sigma_N^{-1} = \sum_{n=1}^{N} x_n \Sigma^{-1} - \mu_0 \Sigma_0^{-1} \tag{23}$$

$$\mu_N = \sigma_N (\sum_{n=1}^{N} x_n \Sigma^{-1} - \mu_0 \Sigma_0^{-1}) \tag{24}$$

$$\mu_N = (\Sigma_0^{-1} - \Sigma^{-1})^{-1} (\sum_{n=1}^{N} x_n \Sigma^{-1} - \mu_0 \Sigma_0^{-1}) \tag{25}$$

2.3 4

$$argmaxp(\mu \mid X, \Sigma, \mu_0, \Sigma_0) = argmaxN(\mu \mid \mu_N, \Sigma_N) \tag{26}$$

We already prove that this is a Gaussian distribution, hence $\mu_{max} = \mu_N = (\Sigma_0^{-1} - \Sigma^{-1})^{-1} (\sum_{n=1}^{N} x_n \Sigma^{-1} - \mu_0 \Sigma_0^{-1})$

3 3

3.1 1

We can assume that each toss x is spread on a bernolli distribution, with the probability μ of being 1 while $1 - \mu$ of being 0.

$$x = \begin{cases} 1 & \mu \\ 0 & 1 - \mu \end{cases} \tag{27}$$

The probability of the forth toss x will therefore become

$$P(x \mid \mu) = \mu_4^x (1 - \mu)^(1 - x_4) \prod_{n=1}^{3} x_n \tag{28}$$

$$argmaxu = \hat{\mu}_{MLE} = \prod_{n=1}^{3} x_n = 1 \tag{29}$$

3.2 2

Now we know the prior follows a beta distribution and the probability of previous data is still μ^3, so the probability of the forth toss will become:

$$p(\mu_4) = \mu^3 \frac{\mu^{(a-1)}(1-\mu)^{(b-1)}}{\beta(a,b)} \tag{30}$$

This still follows a beta distribution. We know that for this distribution, the mode will be the peak of the distribution. Hence the maximum likelihood will be the mode of the distribution. We know that the mode of a $beta(a_1, b_1)$ distribution will be $\frac{a_1-1}{a_1+b_1-2}$. In our example, the $a_1 = 3+a$. Hence the mode will become $\frac{3+a-1}{3+a+b-1} = \frac{2+a}{a+b+1}$

3.3 3

For this question, we know that $\hat{\mu}_{MLE} = \mu^m(1-\mu)^l$ and the prior is still $\frac{\mu^{(a-1)}(1-\mu)^{(b-1)}}{\beta(a,b)}$. Accordingly, the posterior distribution will become

$$p(\mu \mid a,b,m,l) = \mu^m(1-\mu)^l \frac{\mu^{(a-1)}(1-\mu)^{(b-1)}}{\beta(a,b)} \tag{31}$$

$$\propto \mu^{(m+a-1)}(1-\mu)^{(l+b-1)} \tag{32}$$

This is still a beta distribution. The posterior mean will be the mean of beta distribution $\frac{m+a}{m+L+a+b}$. The prior mean is $\frac{a}{a+b}$ and $\hat{\mu}_{MLE} = \frac{m}{m+l}$.

4 4

4.1 1

The exponential family member will always follows this form $log(x \mid \eta) = \eta^T T(x) - A(\eta) + B(x)$

4.1.1 1

The poisson distribution is a member of distribution family

$$pois(l \mid \lambda) = \frac{\lambda^k e^(-\lambda)}{k!} = \frac{exp(log\lambda^k - \lambda)}{k!} \tag{33}$$

Compared with the expression with the exponential family, we could get that

$$T(k) = k \tag{34}$$

$$\eta = log\lambda \tag{35}$$

$$\lambda = exp(\eta) \tag{36}$$

$$A(\eta) = exp(\eta) \tag{37}$$

$$B(k) = -log(k!) \tag{38}$$

4.1.2 2

The Gamma distribution is a member of distribution family

$$Gam(\tau \mid a,b) = \frac{1}{\Gamma(a)} b^a \tau^(a-1)e^(-b\tau) \tag{39}$$

$$= exp(-(-log\frac{1}{\Gamma(a)} - alogb) + (a-1)log\tau - b\tau \tag{40}$$

$$T(\tau) = \tau \tag{41}$$

$$\eta_1 = a-1 \tag{42}$$

$$\eta_1 + 1 = a \tag{43}$$

$$\eta_2 = -b \tag{44}$$

$$A(\eta^T) = -log\frac{1}{\Gamma(\eta_1+1)} - (\eta_1 - 1)log(-\eta_2) \tag{45}$$

$$B(x) = 0 \tag{46}$$

4.1.3 3

$$
\begin{aligned}
Cauchy(x \mid \gamma, \mu) &= \frac{1}{\pi\gamma}\frac{1}{a + (\frac{x-\mu}{\gamma})^2} && (47)\\
&= exp(-log(\pi\gamma) - log(1 + \frac{x^2 - 2\mu x + \mu^2}{\gamma^2})) && (48)
\end{aligned}
$$

The $log(1 + \frac{x^2-2\mu x+\mu^2}{\gamma^2})$ is not separable into the functional multiplication between η and γ. We know this is not a member of exponential family.

4.1.4 4

This distribution is a member of exponential family.

$$
\begin{aligned}
vonMises(x \mid k, \mu) &= \frac{1}{2\pi I_0(k)} exp(kcos(x-\mu)) && (49)\\
&= exp(log\frac{1}{2\pi I_0(k)} + kcos(x-\mu)) && (50)\\
kcos(x-\mu) &= cosxcos\mu + sinxsin\mu && (51)\\
&\begin{cases} kcos\mu = \eta_1 \\ ksin\mu = \eta_2 \end{cases} && (52)\\
\eta_1^2 + \eta_2^2 &= k^2(cos^2\mu + sin^2\mu) && (53)\\
&= k^2 && (54)\\
k &= \sqrt{\eta_1^2 + \eta_2^2} && (55)\\
T(x_1) &= cosx_1 && (56)\\
T(x_2) &= sinx_2 && (57)\\
A(\eta) &= -log\frac{1}{(\sqrt{\eta_1^2 + \eta_2^2}} && (58)\\
B(x) &= log\frac{1}{2\pi I_0} && (59)
\end{aligned}
$$

4.2 2

For exponential family members, we know this rule:

$$
\begin{aligned}
E(T(x)) &= \frac{\partial}{\partial\eta}A(\eta) && (60)\\
Var(T(x)) &= \frac{\partial^2}{\partial^2\eta}A(\eta) && (61)\\
Var(T(x)) &= E[x^2] - (E[x])^2 && (62)
\end{aligned}
$$

We know that for both first and second distribution $x = T(x)$. Hence, we could easily derive that first moment:

$$
\begin{aligned}
1.E(\lambda) = E(T(\lambda)) &= \frac{\partial}{\partial\eta}A(\eta) = exp(\eta) && (63)\\
2.E(\tau) = E(T(\tau)) &= \frac{\partial}{\partial\eta}A(\eta) = -\gamma\gamma(\eta_1 + 1) + \frac{\eta_1 - 1}{\eta_2} && (64)
\end{aligned}
$$

Second moment:

$$
\begin{aligned}
1.Var(\lambda) + E(T(\lambda)) &= E(T(\lambda))^2 && (65)\\
E(T(\lambda))^2 &= exp(2\lambda) + exp(\lambda) && (66)
\end{aligned}
$$

4.3 3

Yes, the poisson distribution has a conjugate prior. The prior of poisson distribution follows a beta distribution and as it is a member of exponential family they its posterior prior follows the same distribution as the beta distribution.