기계인공지능 HW 09 Sol

Problem 1

likelihood $= L(\mu, 6^2) = \prod_{i=1}^{m} c \exp\left(-\frac{1}{26^2}(x_i - \mu)^2\right) = c^m \exp\left(-\frac{1}{26^2}\sum_{i=1}^{m}(x_i-\mu)^2\right)$

Take log. $\Rightarrow$ $m\log C + \left(-\frac{1}{26^2}\sum_{i=1}^{m}(x_i - \mu)^2\right) = \ell(\mu, 6^2)$

To find maximum likelihood w.r.t μ,

differentiate log likelihood w.r.t μ

(since $\log x$ is monotonic & differentiable, max pt does not changed)

Then, $+\frac{1}{6^2}\sum_{i=1}^{m}(x_i - \mu) = 0$ when log likelihood has max value.

So, $\frac{1}{6^2}\sum_{i=1}^{m} x_i - m\mu \cdot \frac{1}{6^2} = 0 \Rightarrow \mu = \frac{1}{m}\sum_{i=1}^{m} x_i$ = sample mean

Problem2

$$\text{Likelihood} = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)$$

$$= \frac{1}{(2\pi)^{\frac{m}{2}}\sigma^m} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i-\mu)^2\right) = L(\mu, \sigma^2)$$

Take log $\Rightarrow$ $-\frac{m}{2}\log 2\pi - m\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i-\mu)^2 = \ell(\mu, \sigma^2)$

To find maximum likelihood w.r.t σ^2,

differentiate log likelihood w.r.t σ^2

(since $\log x$ is monotonic & differentiable, max pt does not changed)

$$-\frac{m}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i-\mu)^2$$

$\Rightarrow$ $-\frac{m}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{m}(x_i-\mu)^2 = 0$ when log likelihood has max value.

Thus, $\sigma^2_{ML} = \frac{1}{m}\sum_{i=1}^{m}(x_i-\mu)^2$

Problem 3

① D: data , θ: parameter

MLE : $\arg\max P(D|\theta)$

MAP : $\arg\max P(\theta|D) = \arg\max P(D|\theta)P(\theta)$

where $P(\theta)$ = prior, $P(D|\theta)$ = likelihood, $P(\theta|D)$ = posterior

② MAP has prior. And it is related to Bayes

$$\frac{P(D|\theta)P(\theta)}{P(D)} = P(\theta|D)$$

③ when $P(\theta)$ follows uniform distribution, $P(\theta)$ is constant.

Then,

$$MAP = \arg\max P(D|\theta)P(\theta) = \arg\max P(D|\theta) = MLE$$

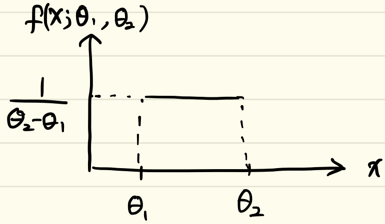
④ same problem with problem 1.

⑤ when prior information is not given, it is the most reasonable to assume it as uniform distribution.

From ③ we can see that MLE assumes prior follows uniform distribution.

Hence, μ_{MLE} assums uniform uniform distribution

Problem 4



$$f(x;\theta_1,\theta_2) = \begin{cases} \frac{1}{\theta_2-\theta_1} & \theta_1 \le x \le \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

If the uniformly distributed r.v are arranged in following order,

$\theta_1 \le x_1 \le \cdots \le x_m \le \theta_2$

The likelihood function is given by

$$L(\theta_1,\theta_2) = \prod_{i=1}^{m} f(x_i) = \left(\frac{1}{\theta_2-\theta_1}\right)^m = (\theta_2-\theta_1)^{-m}$$

Take log $\Rightarrow -m\log(\theta_2-\theta_1) = \ell(\theta_1,\theta_2)$

differentiate log likelihood w.r.t θ_1, θ_2

$$\Rightarrow \frac{\partial \ell}{\partial \theta_1} = \frac{m}{\theta_2-\theta_1} > 0 \quad \text{for } \theta_1 < \theta_2$$

$\therefore \ell$ is increasing for $\theta_1 \Rightarrow \theta_{1MLE} = \min(x_1, \cdots, x_m) = x_1$

$$\Rightarrow \frac{\partial \ell}{\partial \theta_2} = \frac{-m}{\theta_2-\theta_1} < 0 \quad \text{for } \theta_1 < \theta_2$$

$\therefore \ell$ is decreasing for $\theta_2 \Rightarrow \theta_{2MLE} = \max(x_1, \cdots, x_m) = x_m$

Problem5

$$p(x_1, x_2, \dots x_m \mid q) = \prod_{i=1}^{m} q^{x_i} (1-q)^{1-x_i} = L(q)$$

Take log $\Rightarrow$ $\ell(q) = \sum_{i=1}^{m} x_i \log q \quad + \sum_{i=1}^{m} (1-x_i) \log(1-q)$

$$= \log q \sum_{i=1}^{m} x_i + \log(1-q) \sum_{i=1}^{m} (1-x_i)$$

differentiate log likelihood

$$\Rightarrow \frac{\partial \ell}{\partial q} = \frac{1}{q} \sum_{i=1}^{m} x_i - \frac{1}{1-q} \sum_{i=1}^{m} (1-x_i) = 0$$ (when log likelihood has max value.)

$$\Rightarrow (1-q) \sum_{i=1}^{m} x_i - q \sum_{i=1}^{m} (1-x_i) = 0$$

$$\sum_{i=1}^{m} x_i - \cancel{q \sum_{i=1}^{m} x_i} - qm + \cancel{q \sum_{i=1}^{m} x_i} = 0$$

Thus, $q_{ML} = \frac{1}{m} \sum_{i=1}^{m} x_i$

Problem 6

In MLE, we seek a point value for θ which maximizes the likelihood $P(D|\theta)$ and treat term $\frac{P(\theta)}{P(D)}$ as a constant.
($\because$ MLE assume the distribution is uniform)

However, Bayesian estimation fully calculates the posterior distribution $P(\theta|D)$
we have to choose θ that make variance small enough.

In this problem 'SEE' and 'ACT' is a set of observed data and by using data fusion we could get high confidence about our inference.

Problem 7

1) $$\hat{\mu}_{m+1} = \hat{\mu}_m + \frac{1}{m+1}(x_{m+1} - \hat{\mu}_m) \quad \cdots ①$$

$$\hat{\Sigma}_{m+1} = \frac{m}{m+1}\hat{\Sigma}_m + \frac{m}{(m+1)^2}(x_{m+1} - \hat{\mu}_m)(x_{m+1} - \hat{\mu}_m)^T \quad \cdots ②$$

① $$\frac{1}{m}\sum_{i=1}^{m} x_i = \frac{1}{m}\sum_{i=1}^{m} x_i + x_{m+1}$$

$$(m+1)\hat{\mu}_{m+1} = m\hat{\mu}_m + x_{m+1}$$

$$\hat{\mu}_{m+1} = \frac{m+1-1}{m+1}\hat{\mu}_m + \frac{x_{m+1}}{m+1} = \hat{\mu}_m + \frac{1}{m+1}(x_{m+1} - \hat{\mu}_m)$$

② $$\frac{1}{m}\sum_{i=1}^{m}(x_i - \hat{\mu}_{m+1})(x_i - \hat{\mu}_{m+1})^T$$

$$= \sum_{i=1}^{m+1}\left[\left(x_i - \hat{\mu}_m - \frac{1}{m+1}(x_{m+1} - \hat{\mu}_m)\right)\left(x_i - \hat{\mu}_m - \frac{1}{m+1}(x_{m+1} - \hat{\mu}_m)\right)^T\right]$$

$$= \sum_{i=1}^{m}\left[(x_i - \hat{\mu}_m)(x_i - \hat{\mu}_m)^T\right] + \frac{m}{m+1}(x_{m+1} - \hat{\mu}_m)(x_{m+1} - \hat{\mu}_m)^T$$

2) we can estimates more effectively