
Exercise 03 of Machine Learning [IN 2064]

Name: Yiman Li
Matr-Nr: 03724352
cooperate with Kejia Chen(03729686)

Problem 1

Set $f(x) = \theta^t, g(x) = (1 - \theta)^h$, so here $h(x) = f(x)g(x)$, then the first derivative of this likelihood is

$$h'(x) = f'(x)g(x) + f(x)g'(x) = t\theta^{t-1}(1 - \theta)^h - h\theta^t(1 - \theta)^{h-1}$$

and using the same strategy, the second derivative of this likelihood is

$$h''(x) = t(t-1)\theta^{t-2}(1 - \theta)^h - 2t\theta^{t-1}h(1 - \theta)^{h-1} + h(h-1)\theta^t(1 - \theta)^{h-2}$$

Then with respect to the log-likelihood $l(x) = \log\theta^t(1 - \theta)^h = t\log\theta + h\log(1 - \theta)$, the first derivative of this likelihood is:

$$l'(x) = \frac{t}{\theta} - \frac{h}{1 - \theta}$$

and the second derivative of this likelihood is:

$$l''(x) = -\frac{t}{\theta^2} - \frac{h}{(1 - \theta)^2}$$

Problem 2

- The first derivative of the log function is:

$$\frac{d \log f(\theta)}{d\theta} = \frac{1}{f(\theta)} \frac{df(\theta)}{d\theta}$$

Since $f(\theta)$ is positive and differentiable, so

$$\text{when } \frac{df(\theta)}{d\theta} = 0, \text{ we will also have } \frac{d \log f(\theta)}{d\theta} = 0$$

which means that every local maximum of $\log f(\theta)$ is also a local maximum of function $f(\theta)$.

- Since every local maximum of $\log f(\theta)$ is also a local maximum of the differentiable, positive function $f(\theta)$, so we can use $\log f(\theta)$ to simplify the problem when we are going to maximize the likelihood function.

Problem 3

According to the description, we have

$$p(\theta) = \text{Beta}(\theta|6, 4) = \frac{\Gamma(6)\Gamma(4)}{\Gamma(10)}\theta^5(1 - \theta)^3 = \frac{1}{210} \cdot \theta^5(1 - \theta)^3, \theta \in [0, 1]$$

so

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta) \cdot p(\theta)}{p(\mathcal{D})} \propto p(\mathcal{D}|\theta) \cdot p(\theta) \propto \theta^{M+5}(1 - \theta)^{N+3}$$

Then consider the logarithm

$$\theta_{MAP} = \arg \max \log p(\theta|\mathcal{D}) = \arg \max ((M+5)\log\theta + (N+3)\log(1-\theta))$$

Using the first derivative function, we can get:

$$\frac{M+5}{\theta} - \frac{N+3}{1-\theta} = 0$$

That is

$$\theta_{MAP} = \frac{M+5}{M+N+8} = \frac{3}{4}$$

So any (M,N) that can satisfy the equation that $M = 3N + 4$ can lead to such result. For example, $(M, N) = \{(7, 1), (10, 2), (13, 3), \dots\}$

Problem 4

Since a Beta Distribution with parameter a, b has the mean value of $\frac{a}{a+b}$, so the posterior distribution which known as $Beta(\theta|a+m, b+l)$ has the mean value of $\frac{a+m}{a+b+m+l}$, so we need to verify that

$$\frac{a+m}{a+b+m+l} = \frac{a+b}{a+b+m+l} \cdot \frac{a}{a+b} + \frac{m+l}{a+b+m+l} \cdot \frac{m}{m+l}$$

Here, $\lambda = \frac{a+b}{a+b+m+l} \in (0, 1)$, we can figure that

$$\frac{a}{a+b} - \frac{a+m}{a+m+b+l} = \frac{al - bm}{(a+b)(a+m+b+l)} \quad (1)$$

$$\frac{m}{m+l} - \frac{a+m}{a+m+b+l} = \frac{bm - al}{(m+l)(a+m+b+l)} \quad (2)$$

From these two equations we can see that the denominators are always positive, while the numerators of the two equations share the different signs(with one positive and another one negative). Thus we can draw a conclusion that the mean value $\mathbb{E}[\theta|\mathcal{D}]$ of θ lies between the prior mean of and the maximum likelihood estimate for θ .

Problem 5