Machine Learning Worksheet 03

Shang-Hsin Yu – 03681048 – shanghsin.yu@tum.de

Problem 1

The first derivative of original:

$$t\theta^{t-1}(1-\theta)^h - h\theta^t(1-\theta)^{h-1}$$

The second derivative of original:

$$t(t-1)\theta^{t-2}(1-\theta)^h - 2th\theta^{t-1}(i-\theta^{h-1} + h(h-1)\theta^t(1-\theta)^{h-2}$$

The first derivative of log function:

$$\frac{t}{\theta} - \frac{h}{1-\theta}$$

The second derivative of log function:

$$\frac{-t}{\theta^2} - \frac{h}{(1-\theta)^2}$$

Problem 2

Assum that local maximum exists for $f(\theta)$ at $\theta = c$
That means $f'(\theta = c) = 0$ and $f''(\theta = c) < 0$
for any log function of $f(\theta)$, the first derivative of $log(f(\theta))$ is

$$\frac{f'(\theta)}{f(\theta)} = \frac{0}{f(\theta)} = 0$$

which is 0 while $\theta = c$. And the second derivative is

$$\frac{f''(\theta)}{f(\theta)} - \frac{f'(\theta)}{f(\theta)^2} = \frac{f''(\theta)}{f(\theta)} < 0$$

while $\theta = c$. With that we have proven that $log(f(\theta))$ also has local maximum at $\theta = c$
Conclusion is that some derivatives become easier to solve when applied with log

Problem 3

$\theta_{MLE} = \frac{t}{t+h}$ and $\theta_{MAP} = \frac{t+a-1}{t+h+a+b-2}$
the two will be exactly the same if $a = 1$ and $b = 1$, which is the prior that basically states ”I know nothing”

Problem 4

$$p(\theta = x \mid D) = \frac{p(D \mid \theta = x) \cdot p(\theta = x)}{p(D)}$$
$$= \frac{\binom{m+l}{m} x^m (1-x)^l \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}}{p(D)}$$
$$= \frac{x^{m+a-1}(1-x)^{l+b-1}}{Constants}$$

Since this is a probability distribution that looks exactly like a beta distribution, we know the mean of the distribution must be

$$E[\theta \mid D] = \frac{m+a}{m+a+l+b}$$

and we also know the prior mean is beta distribution with parameters a and b, so the mean must be

$$E[\theta] = \frac{a}{a+b}$$

and the θ_{MLE} as known in the lecture is simply

$$\theta_{MLE} = \frac{m}{m+l}$$

finally to prove the result, we choose λ as $\frac{a+b}{m+a+l+b}$, which gives

$$\lambda E[\theta] + (1-\lambda)\theta_{MLE}$$
$$= \frac{a+b}{m+a+l+b} \cdot \frac{a}{a+b} + \frac{m+l}{m+a+l+b} \cdot \frac{m}{m+l}$$
$$= \frac{m+a}{m+a+l+b}$$
$$= E[\theta \mid D]$$

Problem 5

First, we try to compute the λ_{MLE} as follows.
Assume the ith test yields the result t_i, $p(t_1, t_2, t_3, ..., t_n \mid \theta) = \prod_{i=1}^{n} \frac{\lambda^{t_i} e^{-\lambda}}{t_i!}$
we can take the log of this function and get

$$\sum_{i=1}^{n} t_i \log \lambda - \lambda - \log t_i!$$

$$= \log \lambda \cdot \sum_{i=1}^{n} t_i - n\lambda - \sum_{i=1}^{n} t_i!$$

and then we take its first derivative and it yields

$$\sum_{i=1}^{n} t_i \cdot \frac{1}{\lambda} - n$$

we let this equals 0 and get

$$\lambda = \sum_{i=1}^{n} t_i \cdot \frac{1}{n}$$

and then we take the second derivative and get

$$-\sum_{i=1}^{n} t_i \cdot \frac{1}{\lambda^2}$$

which is less than 0, thus we get the local maximum, which is the λ_{MLE}

Next, we plug in the prior knowledge and try to solve for λ_{MAP}

$$p(\lambda \mid D) = \frac{p(D \mid \lambda) \cdot p(\lambda)}{p(D)}$$

$$\propto (\prod_{i=1}^{n} \frac{\lambda^{t_i} e^{-\lambda}}{t_i!}) \cdot \lambda^{\alpha-1} \cdot e^{-\beta\lambda}$$

then we can take its log function

$$-n\lambda + \log x \cdot \sum_{i=1}^{n} t_i - \log t_i! + (\alpha - 1) \log \lambda - \beta\lambda$$

and then we take the first derivative and make it equals to 0

$$-n + \sum_{i=1}^{n} t_i \cdot \frac{1}{\lambda} + \frac{\alpha - 1}{\lambda} - \beta$$

and here we get

$$\lambda = \frac{(\sum_{i=1}^{n} t_i) + \alpha - 1}{n + \beta}$$

in the end, just to make sure, we take the second derivative and get

$$-\frac{(\sum_{i=1}^{n} t_i) + \alpha - 1}{\lambda^2} < 0$$

so what we got is the local maximum and it is the λ_{MAP}