Chaper 3 homework

杨铖章 16337272

2.

(a)

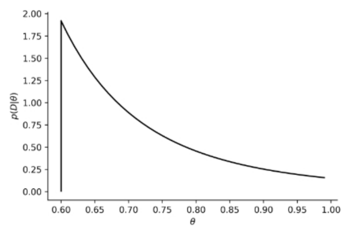
$$L(D|\theta) = \begin{cases} \frac{1}{\theta^n}, \underset{1\le i\le n}{max}\{x_i\} \le \theta \\ 0, otherwise \end{cases}$$

Consider all θ that satisfy $\theta \ge \underset{1\le i\le n}{max}\{x_i\}$, we have

$$L(\theta) = \frac{1}{\theta^n} \le \frac{1}{\underset{1\le i\le n}{max}\{x_i\}}$$

Therefore, $\hat{\theta} = \underset{1\le i\le n}{max}\{x_i\}$

(b)



$$p(D|\theta) = \begin{cases} \frac{1}{\theta^n}, \underset{1\le i\le n}{max}\{x_i\} \le \theta \\ 0, otherwise \end{cases}$$

$p(D^n|\theta)$ is merely relevant to $\underset{1\le i\le n}{max}\{x_i\}$ and the sample number n.

4.

$$\begin{aligned} P(x_1, x_2, \cdots, x_n|\theta) &= \prod_{k=1}^{n}\prod_{i=1}^{d}\theta_i^{x_{ki}}(1-\theta_i^{1-x_{ki}}) \\ L(\theta) &= \sum_{k=1}^{n}\sum_{i=1}^{d}[x_{ki}ln\theta_i + (1-x_{ki})ln(1-\theta_i)] \\ \frac{\partial L}{\partial \theta_i} &= \sum_{k=1}^{n}\sum_{i=1}^{d}(\frac{x_{ki}}{\theta_i} - \frac{1-x_{ki}}{1-\theta_i}) \end{aligned}$$

Let $\frac{\partial L}{\partial \theta_i} = 0$

$$\begin{aligned} (1-\hat{\theta}_i)\sum_{k=1}^{n}x_{ki} &= n\hat{\theta}_i - \hat{\theta}_i\sum_{k=1}^{n}x_{ki} \\ \hat{\theta}_i &= \frac{1}{n}\sum_{k=1}^{n}x_k \end{aligned}$$

21.

When $p(\theta|D^n)$ converge, we have

$$\lim_{n\to\infty} p(\theta|D^n) = \lim_{n\to\infty}\frac{p(x_n|\theta)p(\theta|D^{n-1})}{\int p(x_n|\theta)p(\theta|D^{n-1})d\theta} = \lim_{n\to\infty} p(\theta|D^{n-1})$$

From equation above, we get

$$p(x_n|\theta) = \int p(x_n|\theta)(\theta|D^{n-1})d\theta \tag{1}$$

Assume that

$$\lim_{n\to\infty} p(\theta|D^n) = p(\theta) \ne 0 \tag{2}$$

Also assume that series x_n converges to x^*. To formulate this sentence, we have

$$\lim_{n\to\infty} x_n = x^* \tag{3}$$

From equation(1) to (3), we get

$$\begin{aligned} p(x^*|\theta) &= \lim_{n\to\infty} p(x_n|\theta) \\ &= \lim_{n\to\infty}\int p(x^n|\theta)p(\theta|D^{n-1})d\theta \\ &= \lim_{n\to\infty}\int p(x^n|\theta)p(\theta)d\theta \\ &= \int \lim_{n\to\infty} p(x^n|\theta)p(\theta)d\theta \\ &= \int p(x^*|\theta)p(\theta)d\theta \\ &= p(x^*) \end{aligned}$$

To guarantee the convergence, variables satisfy:

- $\lim_{n\to\infty} p(\theta|D^n) = p(\theta) \ne 0$
- $p(x|\theta) = p(x)$

37.

Recall that

$$\Sigma(\beta) = (1-\beta)\Sigma + \beta I \tag{77}$$

We can see that

$$\lim_{\beta\to 1}\Sigma(\beta) = I$$

Therefore

$$\lim_{\beta \to 1} (1-\beta)\sigma_{ii} + \beta = 1$$

$$\sigma_{ii} = \lim_{\beta \to 1} \frac{1-\beta}{1-\beta} = 1$$

38.

(a)

$$J_1(w) = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} = \frac{w^T(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})^2 w}{w^T(\Sigma_1 + \Sigma_2)w} = \frac{w^T S_B w}{w^T S_W w}$$

We can get the maximun $J_1(w)$ when

$$S_B w = \lambda S_W w$$

Note that S_B is the cross product of two identical vector. S_B always directs to $m_1 - m_2$ where m_1, m_2 refer to the different mean value of two class. In other words, using constant scalar c, we gain
$S_B w = c(m_1 - m_2)$.

Without scalar, we have

$$w = S_W^{-1}(m_1 - m_2) = (\Sigma_1 + \Sigma_2)^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$$

(b)

$$J_2(w) = \frac{(\mu_1 - \mu_2)^2}{P(w_1)\sigma_1^2 + P(w_2)\sigma_2^2} = \frac{w^T(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})^2 w}{w^T(P(w_1)\Sigma_1 + P(w_2)\Sigma_2)w} = \frac{w^T S_B w}{w^T S_W w}$$

Similar to (a), without scalar, we have

$$w = S_W^{-1}(P(w_1)m_1 - P(w_2)m_2) = [P(w_1)\Sigma_1 + P(w_2)\Sigma_2]^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$$

(c)

$$\begin{aligned} J(w) &= \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s_1}^2 + \tilde{s_2}^2} \\ &= \frac{(\mu_1 - \mu_2)^2}{\sum\limits_{y \in D_1}(y-\mu_1)^2 + \sum\limits_{y \in D_2}(y-\mu_2)^2} \\ &= \frac{(\mu_1 - \mu_2)^2}{n(D)(\frac{n(D_1)}{n(D)}(y-\mu_1)^2 + \frac{n(D_2)}{n(D)}(y-\mu_2)^2)} \\ &\approx \frac{(\mu_1 - \mu_2)^2}{n(D)(P(w_1)\sigma_1^2 + P(w_2)\sigma_2^2)} \end{aligned}$$

where $n(D)$ denotes the total number of the sample set. From the result above, we can conclude that (b) is more similar to $J(w)$.

source code

In [1]:

```
def fun2a(x):
    return 1 / x**5 * 1.92 / (1 / 0.6 ** 5)

X = list(np.arange(0.6,1,0.01))
Y = [fun2a(x) for x in X]

ax = plt.gca()

# make right and top spine invisible
for side in ['right','top']:
        ax.spines[side].set_visible(False)
## set cross point
# ax.spines['left'].set_position(('data', 0.5))
# ax.spines['bottom'].set_position(('data', 0))
ax.vlines(0.6, 0, 1.92, colors = 'k');
ax.plot(X, Y, 'k');
plt.xlabel('$\\theta$');
plt.ylabel('$p(D|\\theta)$');

plt.savefig('2a.jpg', dpi = 512);
```