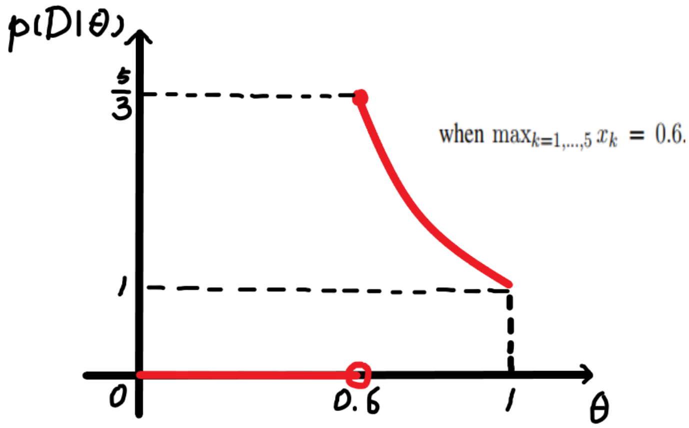
Machine Learning Homework 3

Hairui Yin

1.
(a) Given $max_{k=1,...,5}x_k = 0.6$, the likelihood $p(\mathcal{D}|\theta)$ in $0 \leq \theta \leq 1$ is



Considering maximize the likelihood $p(\mathcal{D}|\theta)$ given 5 samples,

$$
\begin{aligned}
p(\mathcal{D}|\theta) &= \prod_{k=1}^{5} p(x_i|\theta) \\
&= \frac{1}{\theta^5} I(x_1, ..., x_n \in [0, \theta]) \\
&= \frac{1}{\theta^5} I(max(x_1, ..., x_5) \leq \theta)
\end{aligned}
$$

where function I is defined when satisfying the condition then it returns 1, otherwise 0. What it means is that the function I will return 0 when one sample falls outside the interval in $[0, \theta]$.
To maximize the above formula, we need to maximize $\frac{1}{\theta^5}$ given $\theta \geq max(x_1, ..., x_5)$.
Therefore, $\theta_{ML} = max_{k=1,...,5}x_k$.

(b) Considering likelihood $p(\mathcal{D}|\theta)$ given 5 samples,

$$\begin{aligned} p(\mathcal{D}|\theta) &= \prod_{k=1}^{5} p(x_k|\theta) \\ &= \prod_{k=1}^{5} \theta e^{-\theta x_i} I(x_i > 0) \\ &= \prod_{k=1}^{5} \theta e^{-\theta x_i}, \qquad x_i > 0 \\ &= \theta^5 e^{-\theta \sum_{i=1}^{5} x_i} \end{aligned}$$

To maximize $p(\mathcal{D}|\theta)$, we consider its log-likelihood

$$\begin{aligned} l(\theta) &= \log p(\mathcal{D}|\theta) \\ &= 5 \log \theta - \theta \sum_{i=1}^{5} x_i \end{aligned}$$

Since $\sum_{i=1}^{5} x_i > 0$, $5 \log \theta$ is monotonically increasing, $-\theta \sum_{i=1}^{5} x_i$ is monotonically decreasing, it's natural to consider its derivative to find the maximum points.

$$l'(\theta) = \frac{5}{\theta} - \sum_{i=1}^{5} x_i$$

Let $l'(\theta) = 0$, then $\theta = \frac{5}{\sum_{i=1}^{5} x_i}$. To ensure that the point is the maximum point rather than the minimum point, consider its second derivative $l''(\theta) = -5\theta^{-2} < 0$, given $\theta = \frac{5}{\sum_{i=1}^{5} x_i}$. Therefore, the maximum likelihood estimate $\theta_{ML} = \frac{5}{\sum_{i=1}^{5} x_i}$.

2.

(a) The mean vector m (according to Python codes) is

```
mois      40.903067
prot      13.373567
fat       20.229533
ash        2.633233
sodium     0.669400
carb      22.864767
cal        3.271000
```

$$m = [40.90306666666667,\\ 13.373566666666665,\\ 20.229533333333336,\\ 2.633233333333334,\\ 0.6694000000000001,\\ 22.864766666666668,\\ 3.271]$$

(b) The scatter matrix is

	mois	prot	fat	ash	sodium	carb	cal
mois	27286.606579	6620.931419	-4392.180771	963.107725	-108.197648	-30477.235685	-1353.84752
prot	6620.931419	12379.020084	8599.552199	2012.485340	305.765542	-29606.890400	83.80903
fat	-4392.180771	8599.552199	24088.170135	2697.559853	927.666616	-30979.050433	1272.23924
ash	963.107725	2012.485340	2697.559853	482.047164	113.626282	-6153.518024	76.84883
sodium	-108.197648	305.765542	927.666616	113.626282	41.012292	-1238.217242	46.13708
carb	-30477.235685	-29606.890400	-30979.050433	-6153.518024	-1238.217242	97196.196684	-78.49803
cal	-1353.847520	83.809030	1272.239240	76.848830	46.137080	-78.498030	114.94830

To find the first two principal components $\mathbf{w}^1$ and $\mathbf{w}^2$, we need to find two eigenvectors with two largest eigenvalues of the scatter matrix. With python code, the largest two eigenvalues are $\lambda_1 = 126054.97349741 \lambda_2 = 30444.24702927$ and their corresponding eigenvectors (also w_1, w_2) are

$$w_1 = v_1 = [2.76963426e-01, 2.66941457e-01, 2.78933559e-01, 5.54340960e-02,\\ 1.11416057e-02, -8.78084364e-01, 6.03287596e-04]$$

$$w_2 = v_2 = [0.74707368, -0.05573295, -0.65784531, -0.04060421, \\ -0.02381376, 0.00681755, -0.06125383]$$

(c) The value of a_k^l is defined as

$$a_k^l = (x^k - m)^T w_l$$

The first two samples are originally

$$x^1 = [27.82, 21.43, 44.87, 5.11, 1.77, 0.77, 4.93]$$

$$x^2 = [28.49, 21.26, 43.89, 5.34, 1.79, 1.02, 4.84]$$

So the value of $a_1^1, a_1^2, a_2^1, a_2^2$ are

$$a_1^1 = (x^1 - m)^T w^1 = 24.951747905872786$$

$$a_1^2 = (x^1 - m)^T w^2 = -26.811667755251506$$
$$a_2^1 = (x^2 - m)^T w^1 = 24.611975753342847$$
$$a_2^2 = (x^2 - m)^T w^2 = -25.659563399700513$$

Therefore, the approximations $\tilde{x}^1, \tilde{x}^2$ are

$$\begin{aligned}\tilde{x}^1 &= m + a_1^1 w^1 + a_1^2 w^2 \\ &= [27.783496937785824, 21.528515895614635, 44.82734294963899, \\ &\quad 5.105077523179664, 1.5858891661232533, 0.7722370677048186, \\ &\quad 4.928370353002659]\end{aligned}$$

$$\begin{aligned}\tilde{x}^2 &= m + a_2^1 w^1 + a_2^2 w^2 \\ &= [28.550099320064316, 21.373606450091597, 43.97466265199188, \\ &\quad 5.039462273317961, 1.5546676217611726, 1.078440212283421, \\ &\quad 4.8575945711637]\end{aligned}$$

The difference (square error) of x^1 and $\tilde{x}^1$, x^2 and $\tilde{x}^2$ are

$$E_1 = ||x^1 - \tilde{x}^1||^2 = 0.04678617$$

$$E_2 = ||x^2 - \tilde{x}^2||^2 = 0.17311520$$