CSE 6740: Homework 4

Zixuan Wang

November 23, 2023

1 Maximum Likelihood

1.1 Multinomial distribution

Take log of the function f, we can get

$$\begin{aligned}\log f &= \log \frac{n!}{x_1!x_2!\cdots x_k!}\prod_{j=1}^{k}\theta_j^{x_j} \\ &= \log \frac{n!}{x_1!x_2!\cdots x_k!} + \log \prod_{j=1}^{k}\theta_j^{x_j} \\ &= \log \frac{n!}{x_1!x_2!\cdots x_k!} + \sum_{j=1}^{k} x_j \log \theta_j\end{aligned} \tag{1}$$

The constraint is

$$h(\theta) = \sum_{i=1}^{k}\theta_i - 1 = 0$$

The objective function is

$$\max \sum_{j=1}^{k} x_j \log \theta_j + \mu\left(1 - \sum_{j=1}^{k}\theta_j\right)$$

Take the derivative with respect to θ_j and set it to 0,

$$\begin{aligned}\frac{x_j}{\theta_j} - \mu = 0 \\ x_j = \hat{\mu}\cdot\hat{\theta}_j\end{aligned}$$

and $1 - \sum_{j=1}^{k}\theta_j = 0$. Therefore, $\sum_{j=1}^{k} x_j = \hat{\mu}\sum_{j=1}^{k}\theta_j = \hat{\mu} = n$. Thus, $\hat{\theta}_j = \frac{x_j}{n}$.

1.2 Gaussian normal distribution

The log-likelihood function of this distribution is

$$
\begin{aligned}
L &= \log \prod_{i=1}^{m} \left[\frac{1}{|\Sigma|^{1/2}\sqrt{2\pi}^{d}} \exp\left(-\frac{1}{2}(x_i-\mu)^{\top}\Sigma^{-1}(x_i-\mu) \right) \right] \\
&= \sum_{i=1}^{m} -\log(2\pi)^{d/2} - \log|\Sigma|^{1/2} - \frac{1}{2}(x_i-\mu)^{\top}\Sigma^{-1}(x_i-\mu)
\end{aligned}
$$

Take derivate with respect to μ and set it to 0,

$$
-\sum_{i=1}^{m}\Sigma^{-1}\left(x_i-\mu\right) = 0
$$

$$
\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} x_i
$$

Take derivate with respect to Σ and set it to 0,

$$
\hat{\Sigma} = \frac{1}{m}\sum_{i=1}^{m}(x_i-\hat{\mu})(x_i-\hat{\mu})^{\top}
$$

Yes, the ML estimator of Σ is biased.

1.3 Exponential distribution

Given m i.i.d. samples, the log-likelihood function of the distribution is

$$
\begin{aligned}
L &= \log \prod_{i=1}^{m} f(x_i) \\
&= \sum_{i=1}^{m} \log f(x_i) \\
&= \sum_{i=1}^{m} \log \lambda e^{-\lambda x_i} \\
&= m\log\lambda - \lambda\sum_{i=1}^{m} x_i
\end{aligned}
$$

Take the derivative with respect to λ and set it to 0,

$$
\frac{m}{\lambda} - \sum_{i=1}^{m} x_i = 0
$$

$$
\lambda = \frac{m}{\sum_{i=1}^{m} x_i} = \frac{1}{\overline{x}}
$$

2 k-means clustering

2.1

Differentiate J with respect to μ^j and set it to zero,

$$\begin{aligned}\frac{\partial J}{\partial \mu^j} &= \frac{\partial}{\partial \mu^j} \sum_{i=1}^{m} \sum_{j=1}^{k} r^{ij} \|\mathrm{x}_i - \mu^j\|^2 \\ &= \sum_{i=1}^{m} -2r^{ij}(x_i - \mu^j) = 0\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{i=1}^{m} r^{ij} x_i &= \sum_{i=1}^{m} r^{ij} \mu^j \\ \mu^j &= \frac{\sum_i r^{ij} \mathrm{x}_i}{\sum_i r^{ij}}\end{aligned}$$

2.2

Complete linkage.

Single linkage calculates the minimum distance between points in the different clusters, so it is sensitive to outliers and tends to produce long chain-like clusters. Complete linkage calculates the maximum distance and tends to produce compact and spherical clusters that are less sensitive to outliers. Average linkage calculates average distance between all pairs in two clusters, so it tends to produce clusters that are between that produced by single linkage and complete linkage.

k-means tends to produce clusters with similar variance and tries to minimize the maximum distance between points in a cluster. Therefore, complete linkage would most likely result in clusters most similar to those given by k-means.

2.3

None of them would successfully separate the two moons because it is non-convex. Complete linkage might perform slightly better, but it still requires some other techniques.