CSE 6740 A/ISyE 6740, Fall 2023, Georgia Tech

CSE 6740: Homework 4: Unsupervised learning

Due Dec 1st, '23 (11:59 pm ET) on Gradescope

Cite any sources and collaborators; do not copy. See syllabus for policy.

1 Maximum Likelihood [15 pts]

Suppose we have m i.i.d. samples from the following probability distribution. This problem asks you to build a log-likelihood function, and find the maximum likelihood estimator of the parameter(s).

(a) Multinomial distribution [5 pts]

The probability density function of Multinomial distribution is given by

$$f(x_1, x_2, \ldots, x_k; n, \theta_1, \theta_2, \ldots, \theta_k) = \frac{n!}{x_1! x_2! \cdots x_k!} \prod_{j=1}^{k} \theta_j^{x_j},$$

where $\sum_{j=1}^{k} \theta_j = 1, \sum_{j=1}^{k} x_j = n$. What is the maximum likelihood estimator of $\theta_j, j = 1, \ldots k$?

(b) Gaussian normal distribution [5 pts]

Suppose we have m i.i.d. samples from a multivariate Gaussian normal distribution on $\mathbb{R}^d$, $\mathcal{N}(\mu, \Sigma)$, which is given by

$$\mathcal{N}(x; \mu, \Sigma) = \frac{1}{|\Sigma|^{1/2} \sqrt{2\pi}^d} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1} (x-\mu)\right).$$

What is the maximum likelihood estimator of μ and Σ? (4 points) Is the ML estimator of Σ biased? (1 point)

(c) Exponential distribution [5 pts]

The probability density function of Exponential distribution is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

What is the maximum likelihood estimator of λ?

2 k-means clustering [15 pts]

Given m data points $\mathrm{x}_i (i = 1, \ldots, m)$, k-means clustering algorithm groups them into k clusters by minimizing the distortion function over $\{r^{ij}, \mu^j\}$

$$J = \sum_{i=1}^{m} \sum_{j=1}^{k} r^{ij} \|\mathrm{x}_i - \mu^j\|^2,$$

where $r^{ij} = 1$ if x_i belongs to the j-th cluster and $r^{ij} = 0$ otherwise.

Part (a)

Prove that using the squared Euclidean distance $\|\mathrm{x}_i - \mu^j\|^2$ as the dissimilarity function and minimizing the distortion function, we will have

$$\mu^j = \frac{\sum_i r^{ij} \mathrm{x}_i}{\sum_i r^{ij}}.$$

That is, μ^j is the center of j-th cluster. [5 pts]

Part (b)

Suppose at each iteration, we need to find two clusters $\{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_p\}$ and $\{\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_q\}$ with the minimum distance to merge. Some of the most commonly used distance metrics between two clusters are:

- Single linkage: the minimum distance between any pairs of points from the two clusters, i.e.
$$\min_{\substack{i=1,\ldots,p \\ j=1,\ldots,q}} \|\mathrm{x}_i - \mathrm{y}_j\|$$
- Complete linkage: the maximum distance between any parts of points from the two clusters, i.e.
$$\max_{\substack{i=1,\ldots,p \\ j=1,\ldots,q}} \|\mathrm{x}_i - \mathrm{y}_j\|$$
- Average linkage: the average distance between all pair of points from the two clusters, i.e.
$$\frac{1}{mp} \sum_{i=1}^{p} \sum_{j=1}^{q} \|\mathrm{x}_i - \mathrm{y}_j\|$$

Which of the three cluster distance metrics described above would most likely result in clusters most similar to those given by k-means? (Suppose k is a power of 2 in this case). [5 pts]

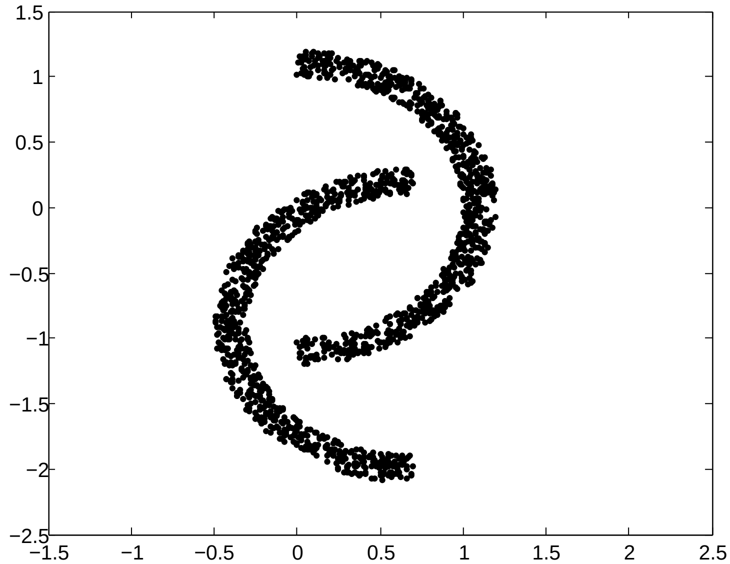

Figure 1: Two moons data

Part c

For the *two moons* data (Figure 1), which of these three distance metrics (if any) would successfully separate the two moons? [5 pts]