STAT 542: Homework 5

Due: April. 5 midnight on Canvas

Please make sure that your solutions are readable and the file size is reasonable. Typing the answers is highly encouraged.

Problem 1.

[2pts] Suppose that $X \in \mathbb{R}^p$ is a random variable taking values of μ_1 and μ_2 with equal probability.

- Find an expression of $\mathbb{E}[XX^\top]$ in terms of μ_1 and μ_2.
- Can a statistician estimate μ_1 and μ_2 (up to permutation) by solving the singular value decomposition of the empirical second moment matrix $\frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top$ (for iid samples $X_1, \ldots, X_n$ following the distribution of X)? Why?
Hint: see p7 of `s13_method_of_moments.pdf`

Problem 2.

Consider 8 data points on the unit circle of the form $(\cos\theta, \sin\theta)$ where

$$\theta = \frac{m\pi}{2} \pm \epsilon, \quad m = 1, 2, 3, 4. \tag{1}$$

Suppose that we want to solve K-means clustering with $K = 4$.

- [2pts] Show that if $\epsilon > 0$ is sufficiently small, the (global) minimum within-point scatter is achieved by pairs of points on the circle with angle difference 2ϵ.
- [1pts] Set $\epsilon = 0.01$, and run Lloyd's algorithm to solve 4-means clustering with random initialization. Run the algorithm 10 times with random initializations and report the minimum value of within-point scatter achieved in the simulation.

  Hint: the R code for kmeans and within-cluster sum of squares is described in p21 of `s12_unsupervised.pdf`.
- [1pts bonus] For sufficiently small $\epsilon > 0$, describe a local minimum of Lloyd's algorithm which is not global optimal, and explain why it has such a property.
- [1pts bonus] For sufficiently small $\epsilon > 0$, describe a local minimum of Lloyd's algorithm which is not a local minimum of Hartigan-Wong, and explain why it has such a property.

Problem 3.

Consider a Gaussian 2-mixture model $P_X = \frac{1}{2}\mathcal{N}(\mu_1, 1) + \frac{1}{2}\mathcal{N}(\mu_2, 1)$. Note that here the weights and the variances of each class are know, so the only unknown parameter is $\theta = (\mu_1, \mu_2) \in \Theta = \mathbb{R}^2$. Let $X_1, \ldots, X_n \in \mathbb{R}$ be observed samples, and $Z_1, \ldots, Z_n \in \{1, 2\}$ be the unobserved class labels.

- [1pts] Give a precise expression of the distribution $p(Z_i = \cdot | X_i, \theta)$ using the logistic function, for any given X_i and θ.
- [1pts] Give an explicit expression of $Q(\beta, \alpha) := \mathbb{E}_{p(Z^n|X^n,\alpha)}[\log p(Z^n, X^n|\beta)]$ in terms of X^n and $\alpha, \beta \in \Theta$, where we defined $Z^n := (Z_1, \ldots, Z_n)$ and $X^n := (X_1, \ldots, X_n)$.
- [1pts bonus] Give an explicit expression of $\arg\max_{\beta \in \Theta} Q(\beta, \alpha)$ in terms of X^n and α.