# CS280 Fall 2018 Assignment 1
# Part A

ML Background

September 13, 2020

**Name:**

**Student ID:**

**1. MLE (5 points)**

Given a dataset $\mathcal{D} = \{x_1, \cdots, x_n\}$. Let $p_{emp}(x)$ be the empirical distribution, i.e., $p_{emp}(x) = \frac{1}{n}\sum_{i=1}^{n} \delta(x, x_i)$ where $\delta(x, a)$ is the Dirac delta function[1] centered at $a$. Assume $q(x|\theta)$ be some probabilistic model.

- Show that $\arg\min_q KL(p_{emp}||q)$ is obtained by $q(x) = q(x; \hat{\theta})$, where $\hat{\theta}$ is the Maximum Likelihood Estimator and $KL(p||q) = \int p(x)(\log p(x) - \log q(x))dx$ is the KL divergence.

[1] `https://en.wikipedia.org/wiki/Dirac_delta_function`

## 2. Gradient descent for fitting GMM (10 points)

Consider the Gaussian mixture model

$$p(\mathbf{x}|\theta) = \sum_k \pi_{k=1}^K \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)$$

Define the log likelihood as

$$l(\theta) = \sum_{n=1}^N \log p(\mathbf{x}_n|\theta)$$

Denote the posterior responsibility that cluster $k$ has for datapoint $n$ as follows:

$$r_{nk} := p(z_n = k|\mathbf{x}_n, \theta) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\mu_k, \Sigma_k)}{\sum_{k'} \pi_{k'} \mathcal{N}(\mathbf{x}_n|\mu_{k'}, \Sigma_k k')}$$

- Show that the gradient of the log-likelihood wrt $\mu_k$ is

$$\frac{d}{d\mu_k} l(\theta) = \sum_n r_{nk} \Sigma_k^{-1} (\mathbf{x}_n - \mu_k)$$

- Derive the gradient of the log-likelihood wrt $\pi_k$ without considering any constraint on $\pi_k$. (bonus 2 points: with constraint $\sum_k \pi_k = 1$.)