Machine Learning Exercise Sheet 12

Clustering

In-class Exercises

K-Medians

Problem 1: Consider a modified version of the K-means objective, where we use L_1 distance instead.

$$\mathcal{J}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\mu}) = \sum_{i=1}^{N} \sum_{k=1}^{K} \boldsymbol{z}_{ik} ||\boldsymbol{x}_i - \boldsymbol{\mu}_k||_1$$

This variation of the algorithm is called K-*medians*. Derive the Lloyd's algorithm for this model.

Gaussian Mixture Model

Problem 2: Derive the E-step update for the Gaussian mixture model.

Problem 3: Derive the M-step update for the Gaussian mixture model.

Expectation Maximization Algorithm

Problem 4: Consider a mixture model where the components are given by independent Bernoulli variables. This is useful when modelling, e.g., binary images, where each of the D dimensions of the image $\boldsymbol{x}$ corresponds to a different pixel that is either black or white. More formally, we have

$$\mathrm{p}(\boldsymbol{x} \mid \boldsymbol{z} = k) = \prod_{d=1}^{D} \boldsymbol{\theta}_{kd}^{x_d} (1 - \boldsymbol{\theta}_{kd})^{1-x_d}.$$

That is, for a given mixture index $\boldsymbol{z} = k$, we have a product of independent Bernoullis, where $\boldsymbol{\theta}_{kd}$ denotes the Bernoulli parameter for component k at pixel d.

Derive the EM algorithm for the parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}_{kd} \mid k = 1, \dots, K, d = 1, \dots, D\}$ of a mixture of Bernoullis.

Assume here for simplicity, that the distribution of components $\mathrm{p}(\boldsymbol{z})$ is uniform: $\mathrm{p}(\boldsymbol{z}) = \prod_{k=1}^{K} \boldsymbol{\pi}_k^{z_k} = \prod_{k=1}^{K} \left(\frac{1}{K}\right)^{z_k}$.

Homework

Gaussian Mixture Model

Problem 5: Consider a mixture of K Gaussians

$$\mathrm{p}(\boldsymbol{x}) = \sum_k \boldsymbol{\pi}_k \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k).$$

Derive the expected value $\mathbb{E}[\boldsymbol{x}]$ and the covariance $\mathrm{Cov}[\boldsymbol{x}]$.

Hint: it is helpful to remember the identity $\mathrm{Cov}[\boldsymbol{x}] = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T] - \mathbb{E}[\boldsymbol{x}]\,\mathbb{E}[\boldsymbol{x}]^T$.

Problem 6: Consider a mixture of K isotropic Gaussians, all with the same *known* covariances $\boldsymbol{\Sigma}_k = \sigma^2 \boldsymbol{I}$.

Derive the EM algorithm for the case when $\sigma^2 \to 0$, and show that it's equivalent to Lloyd's algorithm for K-means.

Problem 7: Consider two random variables $\boldsymbol{x} \in \mathbb{R}^D$ and $\boldsymbol{y} \in \mathbb{R}^D$ distributed according to two different Gaussian mixture models with $\boldsymbol{\theta}^x = \{\boldsymbol{\pi}^x, \boldsymbol{\mu}^x, \boldsymbol{\Sigma}^x\}$ and $\boldsymbol{\theta}^y = \{\boldsymbol{\pi}^y, \boldsymbol{\mu}^y, \boldsymbol{\Sigma}^y\}$, i.e.

$$\mathrm{p}(\boldsymbol{x} \mid \boldsymbol{\theta}^x) = \sum_{k=1}^{K_x} \boldsymbol{\pi}_k^x \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}_k^x, \boldsymbol{\Sigma}_k^x),$$
$$\mathrm{p}(\boldsymbol{y} \mid \boldsymbol{\theta}^y) = \sum_{l=1}^{K_y} \boldsymbol{\pi}_l^y \mathcal{N}(\boldsymbol{y} \mid \boldsymbol{\mu}_l^y, \boldsymbol{\Sigma}_l^y),$$

and the random variable $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$.

a) Describe a generative process (process of drawing samples) for $\boldsymbol{z}$.

b) Explain in a few sentences why $\mathrm{p}(\boldsymbol{z} \mid \boldsymbol{\theta}^x, \boldsymbol{\theta}^y)$ is again a mixture of Gaussians.

c) State the probability density function $\mathrm{p}(\boldsymbol{z} \mid \boldsymbol{\theta}^x, \boldsymbol{\theta}^y)$ of $\boldsymbol{z}$.

Problem 8: Download the notebook `exercise_12_clustering.ipynb` from Moodle. Fill in the missing code and run the notebook. Convert the evaluated notebook to PDF and append it to your other solutions before uploading.