Machine Learning Exercise Sheet 12

Dimensionality Reduction & Clustering

Homework

Matrix Factorization

Problem 1: Download the notebook `exercise_12_matrix_factorization.ipynb` and `exercise_12_matrix_factorization_ratings.npy` from Piazza. Fill in the missing code and run the notebook. Convert the evaluated notebook to PDF and append it to your other solutions before uploading.

Autoencoders

Problem 2: We train a linear autoencoder on D-dimensional data. The autoencoder has a single K-dimensional hidden layer, there are no biases, and all activation functions are identity ($\sigma(x) = x$).

- Why is it usually impossible to get zero reconstruction error in this setting if $K < D$? $W = I$
- Under which conditions is this possible? $K = D$ $D > K$

Gaussian Mixture Model

Problem 3: Consider a mixture of K Gaussians

$\mathbb{E}[x] = p(z \mid X, \theta)$

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

$p(x) = \sum p(x|z) \cdot p(z)$

$p(z|x) = \frac{p(x|z)\, p(z)}{\sum^K p(x|z) \cdot p(z)} = \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}{\sum^K \pi_j \mathcal{N}(x_n | \mu_j, \Sigma_j)} = \gamma(z)$

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 4: 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 $\boldsymbol{z}$.

sample x, sample y $\to z = x + y$

Upload a single PDF file with your homework solution to Moodle by 02.02.2020, 11:59pm CET. We recommend to typeset your solution (using LaTeX or Word), but handwritten solutions are also accepted. If your handwritten solution is illegible, it won't be graded and you waive your right to dispute that.

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

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

Problem 5: Download the notebook `exercise_12_clustering.ipynb` from Piazza. Fill in the missing code and run the notebook. Convert the evaluated notebook to PDF and append it to your other solutions before uploading.

b) because of generative process: x means sample from a gaussian, y aswell => z is a sum of gaussians => also gaussian
But because x and y are mixtures the complete distribution of z is also a gaussian

c) $$p(z) = \sum_k^{K_x} \sum_l^{K_y} \pi_k^x \pi_l^y \mathcal{N}(\mu_k^x + \mu_l^y, \Sigma_k^x + \Sigma_l^y)$$

Problem 3: Law of iterated expectation

$$p(x) = p(x|z) \cdot p(z)$$

$$\mathbb{E}[x] = \mathbb{E}_{p(z)}\Big[\underbrace{\mathbb{E}_{p(x|z)}[x|z]}_{\text{standard gaussian}}\Big]$$

$$= \sum^K \pi_k \mathbb{E}_{p(x|z)}[x|z]$$

$$= \sum^K \pi_k \mu_k$$

$$Cov = \mathbb{E}[xx^T] - \mathbb{E}[x]\mathbb{E}[x]^T$$

$$\mathbb{E}[xx^T] = \mathbb{E}_{p(z)}\left[\mathbb{E}_{p(x|z)}[xx^T|z]\right]$$

$$= \sum^K \pi_k \mathbb{E}_{p(x|z)}[xx^T|z]$$

$$= \sum^K \pi_k (\Sigma + \mu_k \mu_k^T)$$

$$Cov = \sum_k^K \pi_k (\Sigma + \mu_k \mu_k^T) - \sum_k^K \sum_j^K \pi_k \pi_j \mu_k \mu_j^T$$

In-class Exercises

K-Medians

Problem 6: 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 7: Derive the E step update for Gaussian mixture model.

Problem 8: Derive the M step update for Gaussian mixture model.

Expectation Maximization Algorithm

Problem 9: 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 $\boldsymbol{\theta} = \{\boldsymbol{\theta}_{kd} \mid k = 1, \ldots, K, d = 1, \ldots, D\}$ of a mixture of Bernoullis.

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