Machine Learning & Energy - Exercises
8. Bayes' theorem

Tim Janke, Allan Santos

K-means algorithm

Algorithm 1 k-Means

Input: Dataset $\mathbf{X} \in \mathbb{R}^{N \times D}$, initial guess for centroids $\mathbf{c}^0 \in \mathbb{R}^{K \times D}$

Output: optimal centroids $\mathbf{c}^* \in \mathbb{R}^{K \times D}$

$\mathbf{c}^* := \mathbf{c}^0$

do

  $\mathbf{c} := \mathbf{c}^*$

  $\mathbf{z} := \text{find\_closest\_centers}(\mathbf{X}, \mathbf{c})$ $\qquad \triangleright\ \operatorname{argmin}_{k \in \{1 \ldots K\}} ||\mathbf{X} - \mathbf{c}_k||_2^2$

  $\mathbf{c}^* := \text{update\_centroids}(\mathbf{X}, \mathbf{z})$ $\qquad \triangleright\ \frac{\sum_{i \in \{1 \ldots N\}} \mathbb{1}_{\{\mathbf{z}_i = k\}} \mathbf{X}_i}{\sum_{i \in \{1 \ldots N\}} \mathbb{1}_{\{\mathbf{z}_i = k\}}},\ \forall k \in \{1 \ldots K\}$

while $\mathbf{c}^* \neq \mathbf{c}$

EM on Gaussian Mixtures

- **Goal:** Assign data points $\boldsymbol{x}$ to K cluster
  - Differently from k-Means, we want a „soft“ assignment → clusters modeled as Gaussians
  - We have to estimate the Gaussians‘ parameters

- If we knew the cluster z_i of every data point x_i, we could update our estimates for the cluster‘s parameters as:
  - $\mu_k = \frac{\sum_{i=1}^{N} 1_{\{z_i=k\}} x_i}{\sum_{i=1}^{N} 1_{\{z_i=k\}}}$
  - $\Sigma_k = \frac{\sum_{i=1}^{N} 1_{\{z_i=k\}} (x_i - \mu_k)\,(x_i - \mu_k)^T}{\sum_{i=1}^{N} 1_{\{z_i=k\}}}$

- This update is very similar to the one in k-Means, but unfortunately we don‘t know labels (remember this is unsupervised learning).

EM on Gaussian Mixtures

- So, in the Expectation-Maximization (EM) algorithm, instead of using the indicator function, we weight the updates by the likelihood of data point x_i being from cluster k given our current estimates, i.e.:

Expectation step:

- $\Gamma_{i,k} = p(z_i = k \mid x_i, \mu_k, \Sigma_k) = \frac{N(x_i;\mu_k,\Sigma_k)p(z_i=k)}{\sum_{k=1}^{K} N(x_i;\mu_k,\Sigma_k)p(z_i=k)}$, where $z_i \sim Multinomial(\phi)$ is a latent random variable (not observable)

Maximization (of likelihood) step:

- $\mu_k = \frac{\sum_{i=1}^{N} \Gamma_{i,k} x_i}{\sum_{i=1}^{N} \Gamma_{i,k}}$
- $\Sigma_k = \frac{\sum_{i=1}^{N} \Gamma_{i,k}(x_i - \mu_k)\,(x_i - \mu_k)^T}{\sum_{i=1}^{N} \Gamma_{i,k}}$
- $\phi_k = \frac{1}{N}\sum_{i=1}^{N} \Gamma_{i,k}$ → we also need to update the probability of cluster assignment