Problem Set 11 - 12

Computer Vision 2021
University of Bern

1 Expectation Maximization

1. Consider the following data points:
   $x^{(1)} = (1, 1)^T$,
   $x^{(2)} = (1, 3)^T$,
   $x^{(3)} = (2, 1)^T$,
   $x^{(4)} = (2, 3)^T$,
   $x^{(5)} = (7, 1)^T$,
   $x^{(6)} = (7, 3)^T$,
   $x^{(7)} = (8, 1)^T$,
   $x^{(8)} = (8, 3)^T$.
   We start with a hard cluster assignment of the data points, where $p(z^{(i)} = 1 \mid x^{(i)}; \phi, \mu, \Sigma) = 1$ for $i \in \{1, 2, 3, 4\}$ and $p(z^{(i)} = 2 \mid x^{(i)}; \phi, \mu, \Sigma) = 1$ for $i \in \{5, 6, 7, 8\}$. Apply the Expectation Maximisation algorithm to estimate the parameters of the Mixtures of Gaussians model.

2. There is a connection between K-means and the Mixtures of Gaussians model. You can modify the Maximization step of the later by setting $\Sigma = \epsilon \cdot I$, where I is the identity matrix. Prove that when $\epsilon \to 0$, the Expectation Maximization algorithm reduces to the K-means algorithm.

3. Let us denote the dimension of the training data m, and n the number of data points. What happens to the Expectation Maximisation algorithm, when $m > n$?

2 Majorization Minimization

1. **Reweighted least squares.** Consider the linear parameter fitting to the data samples $\{x_i, y_i\}_{i=1}^n$, where $x_i \in \mathcal{R}^N$, $y_i \in \mathcal{R}$ and the objective

function is

$$f(\theta) = \sum_{i=1}^{n} |\theta^T x_i - y_i|. \tag{1}$$

Notice that the model is linear in the parameters $\theta \in \mathcal{R}$, but instead of the usual squared L2-norm we use the L2 norm. This problem does not have an explicit analytical solution, but we can solve it iteratively with MM. Write the formula for the majorizing function $g(\theta|\theta^t)$. Minimize $g(\theta|\theta^t)$ using known methods.
Hint. Use the inequality $2|a||b| \leq |a|^2 + |b|^2$.

2. **Bradley-Terry Ranking**. Consider a sports league with m teams. Assign team i the skill level θ_i. Bradley and Terry proposed the model

$$P(i \text{ beats } j) = \frac{\theta_i}{\theta_i + \theta_j}. \tag{2}$$

To ensure that the skill levels are identifiable, we set $\theta_1 = 1$. If $b_{i,j}$ is the number of times i beats j, then the likelihood of the data is

$$L(\theta) = \prod_{i,j} \left(\frac{\theta_i}{\theta_i + \theta_j} \right)^{b_{i,j}}. \tag{3}$$

Estimate θ by maximizing the log-likelihood $f(\theta) = \log L(\theta)$. Estimate the minorizing function $g(\theta|\theta^t)$ by using the supporting hyperplane. Maximize $g(\theta|\theta^t)$.