Lecture 10: Gaussian Distribution

Course Teacher: Md. Shariful Islam Bhuyan

Likelihood function

- Due to i.i.d assumption, the probability of observing data $\mathbf{d} = x_1, \dots, x_N$

$$p(\mathbf{d}|\theta) = p(x_1, \dots, x_N|\theta) = p(x_1|\theta) \dots p(x_N|\theta) = \prod_{i=1}^{N} p(x_i|\theta)$$

$$L(\theta) = \log p(\mathbf{d}|\theta) = \log p(x_1, \dots, x_N|\theta) = \log \prod_{i=1}^{N} p(x_i|\theta) = \sum_{i=1}^{N} \log p(x_i|\theta)$$

$$\theta_{MLE} = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} \sum_{i=1}^{N} \log p(x_i|\theta) = \arg\min_{\theta} \left(-\sum_{i=1}^{N} \log p(x_i|\theta) \right)$$

$$\theta_{MAP} = \arg\min_{\theta} \left(-\sum_{i=1}^{N} \log p(x_i|\theta) p(\theta) \right) = \arg\min_{\theta} \left(-\sum_{i=1}^{N} \log p(x_i|\theta) - N \log p(\theta) \right)$$

MLE for Univariate Gaussians

- Define likelihood and log-likelihood as objective function

$$l = p(x_1) \dots p(x_N) \quad L = \log p(x_1) \dots p(x_N) = \sum_{i=1}^{N} \log p(x_i) = \sum_{i=1}^{N} \log \frac{1}{\sigma\sqrt{(2\pi)}} e^{\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

$$= \sum_{i=1}^{N} \log \frac{1}{\sigma\sqrt{(2\pi)}} - \sum_{i=1}^{N} \left(\frac{(x_i-\mu)^2}{2\sigma^2}\right) = N\left(-\log\sqrt{(2\pi)} - \log\sigma\right) - \sum_{i=1}^{N} \left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

- Differentiate objective function w.r.t parameters to find analytic solution

$$\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{N} (x_i - \mu) = 0, \qquad \mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$\frac{\partial L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{N} (x_i - \mu)^2 = 0, \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$$

MLE for Multivariate Gaussians [Exercise]

$$p(\mathbf{x}_i|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}} e^{\left(-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})\right)}$$

$$\boldsymbol{\mu} = \frac{\sum_{i=1}^{N}\mathbf{x}_i}{N}$$

$$\boldsymbol{\Sigma} = \frac{\sum_{i=1}^{N}(\mathbf{x}_i-\boldsymbol{\mu})(\mathbf{x}_i-\boldsymbol{\mu})^T}{N}$$