Assignment 1

HongGilDong
Industrial Engineering
UNIST
dmlabisthebest@unist.ac.kr

March 2022

1 K-means algorithm

The k-means algorithm aims to divide a given set of observations into a user-defined number of k clusters. All observations x are allocated to their nearest center-point during each update step (see equation (1)).

$$S_i^{(t)} = \left\{ x_p : \left\| x_p - \mu_i^{(t)} \right\|^2 \leq \left\| x_p - \mu_j^{(t)} \right\|^2 \ \forall j, 1 \leq j \leq k \right\} \tag{1}$$

2 Maximum Likelihood Estimate [10pt]

The likelihood function is nothing but a parameterized density $p(D \mid \boldsymbol{\theta})$ that is used to model a set of data $D = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\}$ which are assumed to be drawn independently from $p(D \mid \boldsymbol{\theta})$:

$$p(D \mid \boldsymbol{\theta}) = p(\boldsymbol{x}_1 \mid \boldsymbol{\theta}) \cdot p(\boldsymbol{x}_2 \mid \boldsymbol{\theta}) \cdot ... \, p(\boldsymbol{x}_n \mid \boldsymbol{\theta}) = \prod_{k=1}^{n} p(\boldsymbol{x}_k \mid \boldsymbol{\theta}) \tag{2}$$

Maximum likelihood seeks to find the optimum values for the parameters by maximizing a likelihood function form the training data. The log-likelihood is given by

$$l(\theta) = \sum_{k=1}^{n} ln \; p(x_k|\theta) \tag{3}$$

3 Gaussian distribution [10pt]

A Gaussian distribution is a type of continuous probability distribution for a real-valued random variable. Univariate Gaussian distribution that is of the form

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right] \tag{4}$$

Multivariate Gaussian distribution that is of the form

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\ |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \right] \tag{5}$$

4 Estimation method derivation [10pt]

Please show the estimation results of maximum likelihood for $p(x) \sim N(\mu|\sigma^2)$, where both μ and σ^2 are unknown.

⊛ Target of the estimation :

μ, σ^2 of Gaussian distribution through point estimation of maximum likelihood estimation.

Our parameterized density $P(X|\theta)$ is given by

$$P(X|\theta) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp\left\{-\frac{1}{2\sigma^2}(X-\mu)^2\right\}$$

Now consider a data set of inputs $\mathcal{X} = \{x_1 \cdots x_N\}$ which are assumed to be independently and identically drawn from $p(X|\theta)$

(i.i.d. assumption)

Let $\theta_1 = \mu$, $\theta_2 = \sigma^2$ are unknown parameters.

We consider the log-likelihood $\mathcal{L}(\theta)$ given by

$$\mathcal{L}(\theta) = \sum_{t=1}^{N} \log P(X_t|\theta) = \sum_{t=1}^{N}\left[-\frac{1}{2}\log 2\pi\theta_2 - \frac{1}{2\theta_2}(x_t-\theta_1)^2\right],$$

To find $\underset{\theta}{\operatorname{argmax}}\,\mathcal{L}(\theta)$, Solve $\nabla_\theta \log P(\mathcal{X}|\theta) = 0$,

(i) Solving $\frac{\partial}{\partial\theta_1}\log P(\mathcal{X}|\theta) = 0$ yields :

$$\frac{\partial}{\partial\theta_1}\mathcal{L}(\theta) = \frac{\partial}{\partial\theta_1}\log P(\mathcal{X}|\theta) = \frac{\partial}{\partial\theta_1}\sum_{t=1}^{N}\left[-\frac{1}{2}\log 2\pi\theta_2 - \frac{1}{2\theta_2}(x_t-\theta_1)^2\right]$$

$$= \sum_{t=1}^{N}\frac{\partial}{\partial\theta_1}\left(-\frac{1}{2\theta_2}(x_t-\theta_1)^2\right) = -\frac{1}{2\theta_2}\sum_{t=1}^{N}\frac{d}{d\theta_1}(x_t-\theta_1)^2$$

$$= -\frac{1}{2\theta_2}\sum_{t=1}^{N} 2\cdot(x_t-\theta_1)\cdot(-1) = \frac{1}{\theta_2}\sum_{t=1}^{N}(x_t-\theta_1) = 0$$

$$\therefore \quad N\cdot\theta_1 = \sum_{t=1}^{N} x_t \quad \Rightarrow \quad \hat{\theta}_1 = \frac{1}{N}\cdot\sum_{t=1}^{N} x_t$$