Inclass 17: Maximum Likelihood Estimate

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.edu)

Department of Artificial Intelligence, Dongguk University

Maximum likelihood estimate

Maximum likelihood estimation, or MLE, is on flavor of parameter estimation in machine learning. In order to perform parameter estimation, we need:

- some data $\mathbf{x}$
- some hypothesized generating function of the data $f(\mathbf{x}, \theta)$
- a set of parameters from that function θ
- some evaluation of the goodness of our parameters (an objective function)

In MLE, the objective function (evaluation) we chose is the likelihood of the data given our model. To find the best θ then, we need to find the θ which maximizes our evaluation function (the likelihood). Therefore, in its general form the MLE is:

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_{\theta} = \arg\max_{\theta} \mathcal{L}(\theta) = \arg\max_{\theta} \mathcal{L}(\theta\,;\mathbf{x}) \tag{1}$$

Maximum likelihood estimate

Gaussian distribution을 따르는 i.i.d. 샘플 $\mathbf{x} = (x_1, x_2, ..., x_N)$로부터 평균 $\theta = \mu$를 MLE로 추정하면,

$$\mathcal{L} = p(\mathbf{x}|\theta) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu) \tag{2}$$

$$\log \mathcal{L} = \sum_{n=1}^{N} \log \mathcal{N}(x_n|\mu) \tag{3}$$

$$\frac{d}{d\mu} \log \mathcal{L} = -const \cdot \sum_{n=1}^{N} (x_n - \mu) \tag{4}$$

$$\frac{d}{d\mu} \log \mathcal{L} = 0 \quad \Longleftrightarrow \quad \hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} x_n \tag{5}$$