

1 Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation (MLE) is a method of estimating the parameters of a statistical model by maximizing the likelihood function.

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \mathcal{L}(\theta; \mathcal{D})$$

$\mathcal{D}$ is our data sample that we observe, $\mathcal{D} = \{y^{(i)}\}_{i=1}^N$.

$\mathcal{L}(\theta)$ is the likelihood of observing our sampled data points for a specific parameter θ , i.e., it is the joint density of the observed data sample given the parameters, $p(y^{(1)}, y^{(2)}, \dots, y^{(N)} \mid \theta)$. $\mathcal{L}$ is a real-valued function of θ so we want to pick the values for our parameters θ that maximizes the likelihood function. With more data points (larger N), we can arrive at parameters that are closer to their true values.

Oftentimes, we take the log of the $\mathcal{L}$ because maximizing the log function is easier. And, maximizing the log function would also maximize the likelihood function.

Very Important!!!: We assume that the data is i.i.d. here

1.1 Multivariate Gaussian

Assume you observe $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$, where each $\mathbf{x}^{(i)} \in \mathbb{R}^M$ is drawn from $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$

Derive the log-likelihood function first.



Now, using this result, derive $\hat{\boldsymbol{\mu}}$. Note that $\frac{\partial}{\partial \mathbf{v}} \mathbf{v}^T A \mathbf{v} = 2A\mathbf{v}$ if A is symmetric