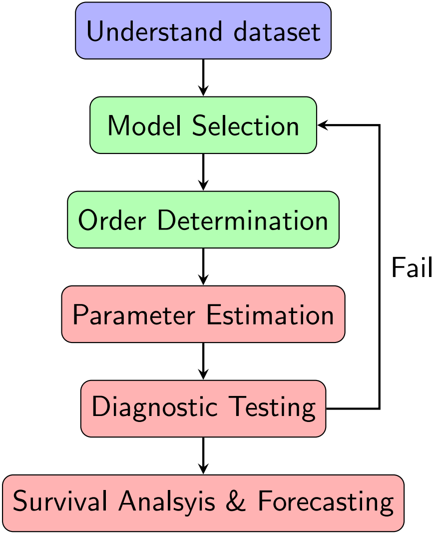
CS5805: Machine Learning I
Lecture : Maximum Likelihood Estimation

Reza Jafari, Ph.D

Collegiate Associate Professor
rjafari@vt.edu

October 22, 2024

Modeling steps



Maximum likelihood Estimation

- There are many methods to estimate unknown parameters from data. One of them is called **Maximum Likelihood Estimation** (MLE).
- For which unknown parameter θ does the observed data $\mathbf{y}$ have the biggest probability?
- MLE is a method of estimating the parameters of probability distribution by maximizing a likelihood function, so that under the assumed statistical model the observed data is most probable.
- The goal of maximize **likelihood estimation** is to find the values of the model parameters that maximize the **likelihood function** over the parameter space.

Bayes Theorem

Setting up Problem

Let θ be the unknown parameter and $\mathbf{y}$ be the set of observations. Using the *Bayes Theorem*:

$$P(\theta|\mathbf{y}) = \frac{P(\mathbf{y}|\theta)P(\theta)}{P(\mathbf{y})}$$

$$posterior = \frac{Likelihood \times prior}{evidence}$$

Basic Idea of MLE

- Suppose we have a random samples $(y_1,\ y_2,\ldots,y_T)$ whose assumed probability distribution depends on some unknown parameter θ.
- It seems reasonable that a good estimate of the unknown parameter θ would be the value of θ that **maximizes** the probability $P(\mathbf{y}|\theta)$, the **likelihood**, of getting the data we observed.
- In a nutshell, that is the idea behind the method of maximum likelihood estimation.
- MLE is the most general estimator for parameter estimation.

How to implement MLE

1. Let suppose the joint probability density function of the measurements are given as :

$$f_{Y_1,Y_2,\ldots,Y_T}(\lambda_1, \lambda_2, \ldots, \lambda_T; \theta) = f_{Y_1}(\lambda_1; \theta).f_{Y_2}(\lambda_2; \theta), \ldots, f_{Y_T}(\lambda_T; \theta)$$

2. Replace dummy variable by the measurements to construct likelihood function (function of θ):

$$L(\mathbf{y}|\theta) = f_{Y_1}(y_1; \theta).f_{Y_2}(y_2; \theta), \ldots, f_{Y_T}(y_T; \theta)$$

3. For simplifying calculations, it is customary to work with the natural logarithm of L as the likelihood function. Hence, $\hat{\theta}_{MLE}$ calculated as :

$$\boxed{\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\, L(\mathbf{y}|\theta) \quad or \quad \underset{\theta}{\operatorname{argmax}}\, \ln(L(\mathbf{y}|\theta))}$$

MLE

- If $\ln(L(\mathbf{y}|\theta))$ is used to find the θ_{MLE}, this is called **log-likelihood**.
- In order to find θ_{MLE}, the necessary condition is :

$$\frac{\partial L(\mathbf{y}|\theta)}{\partial \theta} = 0$$

which is called likelihood equation. $\hat{\theta}_{MLE}$ is the solution to above equation.

Example

Suppose our data $y_1, ..., y_T$ are independently drawn from a uniform distribution $U(a, b)$. Find the MLE of $\hat{a}$,$\hat{b}$.

Example

Suppose that we have observed random samples $y_1, y_2, ..., y_N$ where $y_i \sim N(\theta_1, \theta_2)$. Find the maximum likelihood estimator for θ_1, θ_2

$$f_y(\lambda) = \frac{1}{\theta_2\sqrt{2\pi}} e^{\frac{-(\lambda-\theta_1)^2}{2\theta_2^2}}$$