Maximum Likelihood Parameter Estimation

Ghazal Farhani

June 27, 2020

1 Max likelihood in the case of Bernoulli

The Likelihood of a sequence $D = x_1, x_2, ..., x_n$ (of for example coin tosses) can be written as:

$$
\begin{aligned}
P(D|\theta) &= \Pi_{n=1}^{N} \theta_n^x (1-\theta)^{1-x_n} = \theta_1^N (1-\theta)^{N_0} \\
N_1 &= \Sigma_{n=1}^{N} x_n \\
N_0 &= \Sigma_{n=1}^{N} (1-x_n)
\end{aligned}
\tag{1}
$$

In stead of likelihood we can find log of likelihood. The maximum log likelihood $\hat{\theta} = argmaxP(D|\theta)$ for the Bernoulli's distribution can be calculated as follows:

$$
\begin{aligned}
l(\theta) = logP(D|\theta) &= log\theta_1^N (1-\theta)^{N_0} \\
&= N_1 log\theta + N_0 log(1-\theta)
\end{aligned}
\tag{2}
$$

Knowing $l(\theta)$, to calculate $\hat{\theta}$ the gradient of $l(\theta)$ with respect to θ should be found:

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{N_1}{\theta} - \frac{N_0}{1-\theta} \tag{3}$$

And $\frac{\partial l(\theta)}{\partial \theta} = 0$ will result in:

$\hat{\theta} = \frac{N_1}{N}$

2 Bayesian Parameter Estimation

Instead of having a deterministic approach to estimate the maximum value for θ a probabilistic option can be considered based on a conditional probability approach:

$$
\begin{aligned}
P(D|\theta) &= \frac{P(\theta|D)P(\theta)}{P(D)} \\
P(D) &= \int P(D|\theta) d\theta
\end{aligned}
\tag{4}
$$

$P(D)$ is the normalization factor which is independent of θ and is called marginal likelihood evidence, and $P(\theta)$ is the *a priori*.

2.1 Conjugate Prior

A *prior* is called conjugate if the result of multiplication of *prior* and $P(D|\theta)$ be a posterior with the same parametric family of *prior*. In the case of Bernoulli: