Maximum Likelihood Estimation

Wednesday 1 May 2024 4:55 PM

1) Random Experiment → Distribution → Parameters → event chance [probability]

2) [Likelihood] questioning a parameter's validity ← observed event/data

→ probability is a measure of the chance that a certain event will occur out of all possible events.

→ Likelihood is a function that measures the plausibility of a particular parameter given some observed data. it quantifies how well a specific outcome supports specific parameter values. i.e. given observed data how justified are a specific set of parameter values.

→ Maximum likelihood Estimation (MLE) is a method of estimating the parameters of a statistical model given some observed data. i.e. given the observed data, find parameters which are most likely/most justifiable/which are supported by the observed data the most/which maximize the likelihood function

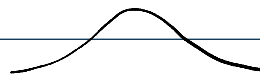
"Best model given the observed data"

MAX [L (parameters|observed data)]

Sample data → Assume distribution → MLE → Find the value of parameter(s) that maximize the likelihood function.

MLE for Normal Distribution

Observed data = $\{x_1 \; x_2 \; x_3 \; \dots \; x_n\}$



1

QQ plot, shapiro wilk test, histogram/kde

which $N(\mu, \sigma)$??

→ Based on the observed data find μ, σ for which likelihood function is maximum.

3 $MAX[L(\mu, \sigma | x_1, x_2, \dots x_n)]$ 2 $L(\mu, \sigma | x_1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}$

$$\log(L) = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{(x_1-\mu)^2}{2\sigma^2} - \frac{(x_2-\mu)^2}{2\sigma^2} - \dots - \frac{(x_n-\mu)^2}{2\sigma^2}$$

$$\frac{\partial L}{\partial \mu} = \frac{(x_1-\mu)+(x_2-\mu)+\dots+(x_n-\mu)}{\sigma^2} = 0$$

$$\sum_{i=1}^{n}(x_i) - n\mu = 0$$

$$\boxed{\mu = \frac{\sum_{i=1}^{n}(x_i)}{n}}$$

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{(x_1-\mu)^2}{\sigma^3} + \frac{(x_2-\mu)^2}{\sigma^3} + \dots + \frac{(x_n-\mu)^2}{\sigma^3} = 0$$

$$\frac{(x_1-\mu)^2}{\sigma^3} + \frac{(x_2-\mu)^2}{\sigma^3} + \dots + \frac{(x_n-\mu)^2}{\sigma^3} = \frac{n}{\sigma}$$

$$\frac{(x_1-\mu)^2}{\sigma^2} + \frac{(x_2-\mu)^2}{\sigma^2} + \dots + \frac{(x_n-\mu)^2}{\sigma^2} = n$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$$

$$\boxed{\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}}}$$

MLE in Machine Learning

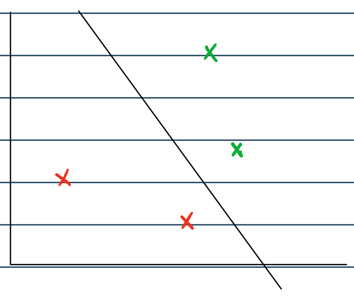
x_1	x_2	x_3	x_4	y

1) find out distribution of $y|x$
2) decide to apply a ml model that is parametric in nature
3) initiate model parameters with random values
4) select a likelihood function
5) find model parameters for which the likelihood function is maximum.

→ MLE can be used to train a ml model given it's parametric.

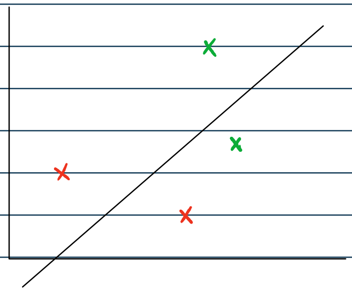
MLE in logistic Regression

A



OR

B



x_1	x_2	x_3	x_4	y
				1
				0
				1
				1

→ bernoulli distribution ①

→ Logistic regression ②

↓

Random values for ③ $\beta_0\ \beta_1 \dots\dots \beta_n$

↓

$p^y(1-p)^{(1-y)}$ ④

Assume y_i's are independent

$$L(y|x;\beta) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{(1-y_i)}$$

$$\log(L) = \sum_{i=1}^{n} y_i \log(p_i) + (1-y_i)\log(1-p_i)$$

→ maximize

Apply gradient descent to find β for which $\log(L)$ is maximum

Questions

1) is MLE applicable to all ML algos? NO
- Non-parametric models X
- Unsupervised learning algorithms X
- Reinforcement learning X

2) Relation of Loss function & MLE?

→ MLE and the concept of loss function in ML are closely related. Many common loss functions can be derived from the principle of MLE under certain assumptions about the data/model. By minimizing the loss function we're effectively performing MLE.

3) Purpose of loss function when we have MLE?

1) Computational reason
2) Generalization
3) Flexibility

TRY

1) Likelihood function of softmax regression & compare with categorical loss entropy.

2) MLE to linear regression $y - \hat{y} \rightarrow N(\mu, \sigma)$