Maximum Likelihood Estimation

Aakash Daswani Aayush Arora

Intro To AI / ML , February 2019

Table of Contents

1. Introduction
2. Intutive Explanation of MLE

Table of Contents

1. Introduction
2. Intutive Explanation of MLE

Introduction

What are Parameters ?

- Often in machine learning we use a model to describe the process that results in the data that are observed.
- Each model contains its own set of parameters that ultimately defines what the model looks like.
- For a linear model we can write this as $y = mx + c$, m and c are parameters for this model. Different values for these parameters will give different lines
- parameters define a blueprint for the model. It is only when specific values are chosen for the parameters that we get an instantiation for the model that describes a given phenomenon

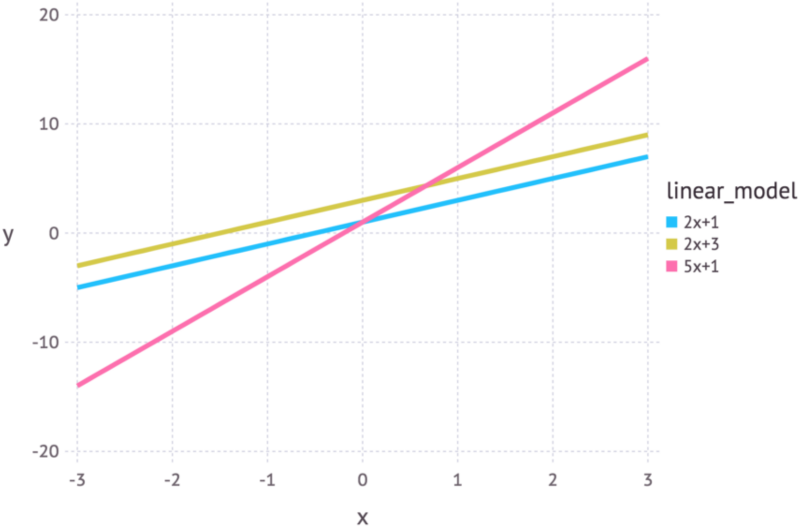

20
10
0
-10
-20
y
-3
-2
-1
0
1
2
3
x
linear_model
2x+1
2x+3
5x+1

Table of Contents

1 Introduction

2 Intutive Explanation of MLE

Intutive Explanation of MLE

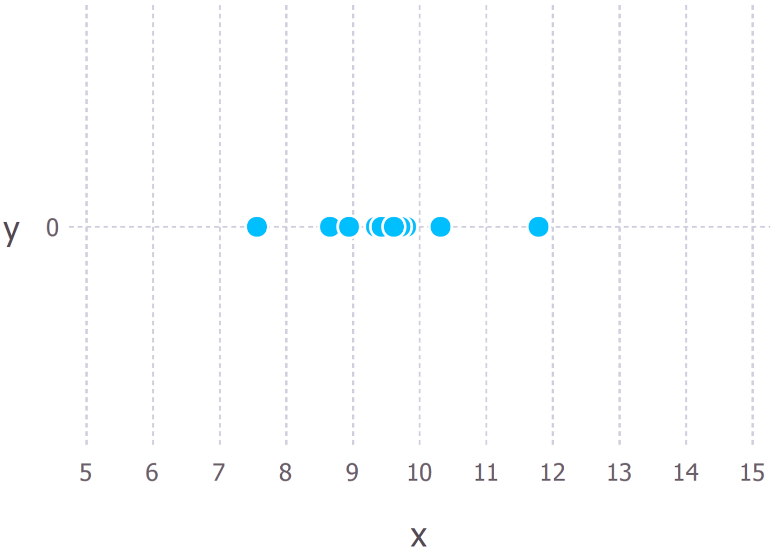
- likelihood estimation is a method that determines values for the parameters of a model.
- parameter values are found such that they maximise the likelihood that the process described by the model produced the data that were actually observed.

Three Major Steps in Applying MLE

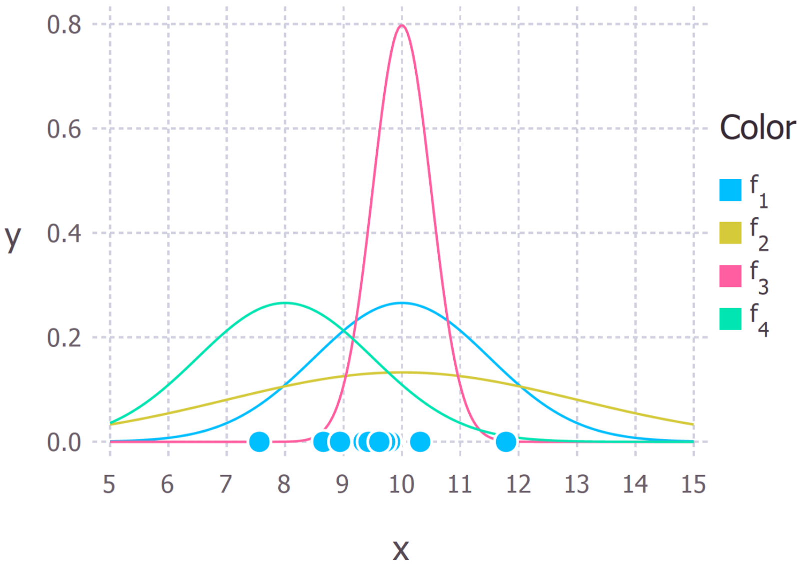
- Define the likelihood, ensuring you're using the correct distribution for your regression or classification problem
- Take the natural log and reduce the product function to a sum function.
- Maximize or minimize the negative of the objective function

Example

- Let's suppose we have observed 10 data points from some process. For example, each data point could represent the length of time in seconds that it takes a student to answer a specific exam question. These 10 data points are shown in the figure below


y
0
5
6
7
8
9
10
11
12
13
14
15
x

- we'll assume that the data generation process can be adequately described by a Gaussian (normal) distribution
- The Gaussian distribution has 2 parameters. The mean and the standard deviation, Different values of these parameters result in different curves
- We want to know which curve was most likely responsible for creating the data points we observed
- We'll use MLE to estimate these parameters


0.8
0.6
0.4
0.2
0.0
y
5
6
7
8
9
10
11
12
13
14
15
x
Color
f_1
f_2
f_3
f_4

Calculating MLE

- Suppose we have three data points this time and we assume that they have been generated from a process that is adequately described by a Gaussian distribution. These points are 9, 9.5 and 11
- *How do we calculate the maximum likelihood estimates of the parameter values of the Gaussian distribution ?*

The probability density of observing a single data point x, that is generated from a Gaussian distribution is given by:

$$P(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

In our example the total (joint) probability density of observing the three data points is given by:

$$P(9, 9.5, 11; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(9-\mu)^2}{2\sigma^2}\right) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(9.5-\mu)^2}{2\sigma^2}\right) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(11-\mu)^2}{2\sigma^2}\right)$$

log likelihood

Taking logs of the original expression gives us:

$$\ln(P(x;\mu,\sigma)) = \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(9-\mu)^2}{2\sigma^2} + \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(9.5-\mu)^2}{2\sigma^2} + \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(11-\mu)^2}{2\sigma^2}$$

$$\ln(P(x;\mu,\sigma)) = -3\ln(\sigma) - \frac{3}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\left[(9-\mu)^2 + (9.5-\mu)^2 + (11-\mu)^2\right]$$

To do this we take the partial derivative of the function with respect to μ, giving

$$\frac{\partial \ln(P(x;\mu,\sigma))}{\partial \mu} = \frac{1}{\sigma^2}\left[9 + 9.5 + 11 - 3\mu\right].$$

which gives $\mu = 9.833$
Differentiating with respect to σ gives :
$\sigma = 0.848$