Maximum Likelihood Estimation

The Logistic Regression is built on **Maximum Likelihood Estimation**

Probability vs Likelihood

- In probability we know the *parameters* which exactly describe the situation and how often things occur (Something must happen), they add up to **one** everything happen on the same universe
  For a single value of parameter something must happen
- Likelihoods are Probability of the **Observed data** under a hypothetical scenario.
- There are many likelihoods that do not add up to one and thus cannot be interpreted as probabilities
- **Likelihoods** depends on the parameter
  Given this **Observed data** what parameter θ make it probable, explains the data **Observed**
  - The **Maximum Likelihood** chooses the universe where are data would be most likely

Example :

- Flipping a coin one time and Observing **heads**

Consider these **universes**

1. **Probability** of heads is very small
2. **Probability** of heads or tails is fair
3. **Probability** of heads is 100%
   The **Maximum Likelihood** is the Probability of a coin that always lands on **heads** we maximize the chances of landing hands

- The probability of the data we observed is maximized
  But most of the time we add restrictions and study the probability of that happening with unknown parameters
- Probability of heads is p
- Probability of tails is $1-p$
- The goal is to estimate what p that maximizes the likelihood

- This solved using **Derivations** and studying where the graph maximize

p	1-p	Likelihood $p^3 \times (1-p)^2$
0.0	1.0	0.00000
0.1	0.9	0.00081
0.2	0.8	0.00512
0.3	0.7	0.01323
0.4	0.6	0.02304
0.5	0.5	0.03125
0.6	0.4	0.03456
0.7	0.3	0.03087
0.8	0.2	0.02048
0.9	0.1	0.00729
1.0	0.0	0.00000

- in a **Normal distribution** Setting the likelihood is maximize when we choose the **mean** parameter that best fit the data. The **MLE** of the mean of a normal distribution is the **Sample mean**. Searching for the value of the mean that makes the observed data most probable under the normal distribution
- If we have two observed probabilities the **MLE** try to maximize both of the probabilities