MLE Notes

Wednesday, July 1, 2020 7:52 AM

what is the probability of flipping Heads? In the absence of knowing the coin is fair...

$P(H) = 3/5$

Flip 5 times

H T H H T

Flip 50 times

$H = 30,\ T = 20 \Rightarrow P(H) = 3/5$

Maximum Likelihood Estimation (MLE)

Dataset $D = (H\,T\,H\,H\,T\,\ldots)$

Hypothesis: $P(H) = \theta$
$P(T) = 1-\theta$] assumption of how the world works

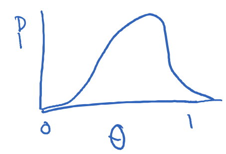
Assumption:

Data i.i.d. → independent, identically distributed

$P(HTHHT) = P(H)\,P(T)\,P(H)\,P(H)\,P(T)$

$P(D|\theta) \quad = \theta^3 (1-\theta)^2$

data given θ



$k \Rightarrow$ heads
$n \Rightarrow$ flips

$P(D|\theta) = \theta^k (1-\theta)^{n-k}$

find θ that maximizes this probability

$$\hat{\theta}_{MLE} = \max_\theta P(D|\theta)$$
$$= \max_\theta \theta^k (1-\theta)^{n-k}$$
$$= \max_\theta \log\left(\theta^k (1-\theta)^{n-k}\right)$$
$$= \max_\theta k\log\theta + (n-k)\log(1-\theta)$$

$\log(xy) = \log(x) + \log(y)$

$\log(x^y) = y\log(x)$

Bayes' rule

$$P(\theta|D) = \frac{P(D|\theta)\,\cancel{P(\theta)}}{\cancel{P(D)}}$$

every θ is equally likely

independent of θ

$$\frac{d}{d\theta}\log P(D|\theta) = \ldots$$

$$\hat{\theta}_{MLE} = \frac{k}{n}$$ ← # heads, ← # flips

- unbiased
- approaches true relationships

Bayesianists → should take into account assumptions

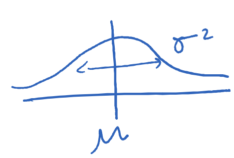
Frequentists → look at the data

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

mean mu

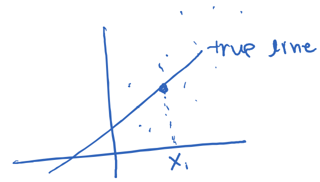
variance sigma squared

continuous. prob. dist



Linear Regression

$y_i = W^T x_i + \varepsilon_i$ $\quad\quad$ $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$



$P(y_i | x_i) \sim \mathcal{N}(W^T x_i, \sigma^2)$

$D = ((x_i, y_i) \ldots (x_n, y_n))$

$\hat{W}_{MLE} = \max_W P(D | W)$ $\quad$ find the w that makes my data the most likely

$= \max_W P(y_1 | x_1, W)\, P(y_2 | x_2, W)\, P(y_3 | x_3, W) \ldots P(y_n | x_n, W)$

$= \max_W \sum_{i=1}^{n} \log\left(P(y_i | x_i, W)\right)$

$= \max_W \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - W^T x_i)^2}{2\sigma^2}}\right)$

$\ldots$

$= \max_w - \sum_{i=1}^{n} (y_i - W^T x_i)^2$

$\hat{w}_{MLE} = \min_w \sum_{i=1}^{n} (y_i - W^T x_i)^2$ $\quad$ most likely value of w is the one that minimizes RSS