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

\[ P(H) = \frac{3}{5} \]

Flip 5 times
HTHTT
Flip 50 times
H=30, T=20 \Rightarrow P(H) = \frac{3}{5}

**Maximum Likelihood Estimation (MLE)**

Dataset \( D = (HTHTT, \ldots) \)

Hypothesis:
\[ P(H) = \theta \quad \text{and} \quad P(T) = 1 - \theta \]

Assumption:
Data i.i.d. \( \rightarrow \) independent, identically distributed

\[ P(D | \theta) = P(H)^k P(T)^{n-k} \]

Find \( \theta \) that maximizes this probability

\[ \hat{\theta}_{MLE} = \max_{\theta} P(D | \theta) \]

\[ = \max_{\theta} \theta^k (1-\theta)^{n-k} \]

\[ = \max_{\theta} \log(\theta^k (1-\theta)^{n-k}) \]

\[ = \max_{\theta} k \log(\theta) + (n-k) \log(1-\theta) \]

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

\[ \hat{\theta}_{MLE} = \frac{k}{n} \text{ heads} \]

\[ \log(x+y) = \log(x) + \log(y) \]

\[ \log(x^n) = n \log(x) \]

Bayes' rule:

\[ P(\theta | D) = \frac{P(D | \theta) \cdot P(\theta)}{P(D)} \]

\( P(D) \) independent of \( \theta \)

Bayesianists \( \rightarrow \) should take into account assumptions.
Frequentists \( \rightarrow \) look at the data.

\[ N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
Linear Regression

\[ y_i = w^T x_i + \epsilon_i \]
\[ \epsilon_i \sim \mathcal{N}(0, \sigma^2) \]

\[ P(y_i \mid x_i) \sim \mathcal{N}(w^T x_i, \sigma^2) \]

\[ D = (x_1, y_1), \ldots, (x_n, y_n) \]

\[ \hat{w}_{\text{MLT}} = \max_w P(D \mid w) \quad \text{find the } w \text{ that makes my data the most likely} \]
\[ = \max_w P(y_1 \mid x_1, w) P(y_2 \mid x_2, w) P(y_3 \mid x_3, w) \ldots P(y_n \mid x_n, w) \]
\[ = \max_w \sum_{i=1}^{n} \log P(y_i \mid x_i, w) \]
\[ = \max_w \sum_{i=1}^{n} \log \left( \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - w^T x_i)^2}{2\sigma^2}} \right) \]
\[ = \max_w - \sum_{i=1}^{n} (y_i - w^T x_i)^2 \]

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