Implementing

Naive Bayes classifier

- Bayes Theorem : $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$

- In our case: $P(y|X) = \frac{P(X|y) \cdot P(y)}{P(X)}$

  - with feature vector X : $X = (x_1, x_2, x_3, \dots, x_n)$

- Assume that all features are mutually independent

$$P(y|X) = \frac{P(x_1|y) \cdot P(x_2|y) \cdot \dots \cdot P(x_n|y) \cdot P(y)}{P(X)}$$

- use the chain rule so we calculate the probability for each feature
- $P(y|X)$ is the posterior probability
- $P(x_i|y)$ is the class conditional probability
- $P(y)$ is the prior probability of y
- $P(X)$ is the prior probability of X

- Select class with highest probability

$$y = \text{argmax}_y \, P(y|X) = \text{argmax}_y \frac{P(x_1|y) \cdot P(x_2|y) \cdot \dots \cdot P(x_n|y) \cdot P(y)}{P(X)}$$

$$y = \text{argmax}_y \, P(x_1|y) \cdot P(x_2|y) \cdot \dots \cdot P(x_n|y) \cdot P(y)$$

- since we are only interested in why we don't need $P(X)$

$$y = \text{argmax}_y \log(P(x_1|y)) + \log(P(x_2|y)) + \dots + \log(P(x_n|y)) + \log(P(y))$$

- Class conditional probability $P(x_i|y)$
  - we model this with a Gaussian distribution

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \cdot \exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)$$