Week 6

Naive Bayes Classifier

> 💡 sklearn.naive_bayes

For a given class variable and dependent feature vector through, the naive conditional independence assumption is given by:

$$P(x_i \mid y, x_1, ..., x+i-1, x_{i+1}, ..., x_m) = P(x_i \mid y)$$

Types

- **Gaussian NB:** continous features

- **Bernoulli NB:** binary features

- **Multinomial NB:** multinomial distribution, classification of word counts for text classification

- **Categorical NB:** categorical distribution

- **Complement NB:** class imbalance, CNB regularly outperforms MNB (often by a considerable margin) on text classification tasks.