Patern Recognition Principles Week6 Assignment

#### **1. Conditional Independence Assumption in Document Classification**

In the context of document classification, the **Naive Bayes algorithm** assumes that all the features (i.e., words in a document) are **conditionally independent of each other given the class label**.

Formally, for a document ( D = {w\_1, w\_2, …, w\_n} ) and class ( C ):

This means that once we know the class (e.g., “sports” or “politics”), the occurrence of one word in the document does not affect the probability of another word appearing.

#### **2. Importance and Computational Benefits**

This “naive” assumption is crucial because it makes the computation **tractable**. Without it, estimating ( P(w\_1, w\_2, …, w\_n | C) ) would require modeling the joint probability of all possible word combinations — an enormous number that grows exponentially with the number of features (the **curse of dimensionality**).

By assuming independence, Naive Bayes reduces the parameter estimation from exponential ( O(k^n) ) (where ( k ) is the number of word probabilities per feature) to linear ( O(nk) ). This drastically simplifies learning and allows the model to perform well even with relatively small amounts of training data.