Assignment

**Question 1.** *In the context of classifying a document, what is the conditional independence assumption made by the Naive Bayes algorithm regarding the words (features) in the document?*

Naive Bayes treats every word position in a document as a separate feature and assumes that, once we know the true class label , the probability of any word appearing is **independent** of the presence or absence of every other word. Formally, for a document represented by the vector of words , the model posits:

Hence, the joint likelihood of the entire feature vector factorises into a product of individual word likelihoods, each conditioned only on the class.

**Question 2.** *Explain why this “naive” assumption is so useful and how it helps to overcome the “curse of dimensionality” by reducing the number of parameters the model needs to estimate.*

Without the independence assumption, estimating would require storing a probability for **every possible combination of words**—exponential in the vocabulary size.  
By factorising the joint distribution, Naive Bayes reduces the task to learning one parameter per word per class, i.e. . For a vocabulary of size and classes, the total number of parameters drops from to . This linear scaling keeps memory and computational demands feasible even when thousands or tens of thousands of distinct features (words) are present, thereby taming the curse of dimensionality and making the classifier both trainable and fast at prediction time.