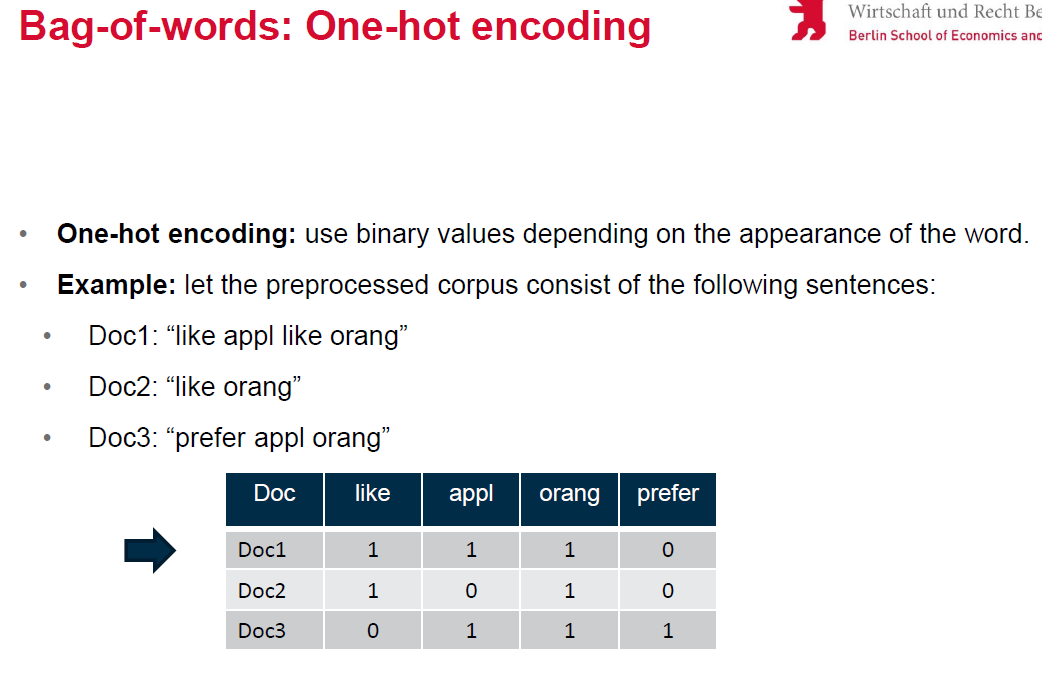
2nd Lesson

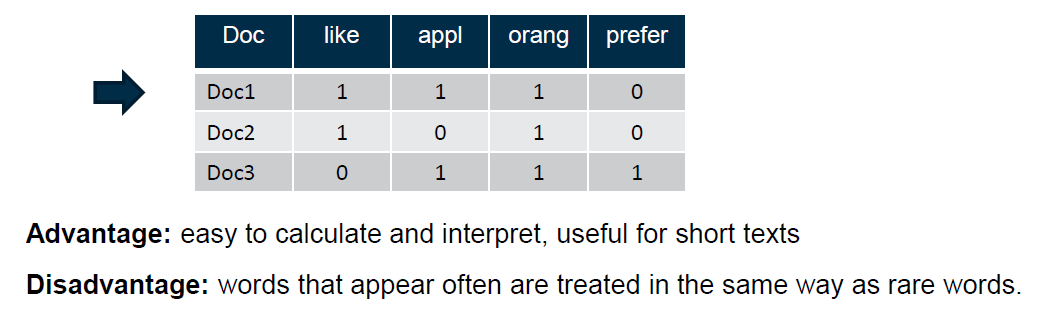
Weakness of Sentiment analysis: negations

* One hot encoding
* Absolut Term Frequencies
* Relative Term Frequencies
* Weighted Term Frequencies

TF 🡪 Document Frequencies

# Bag-of-words: One-hot encoding

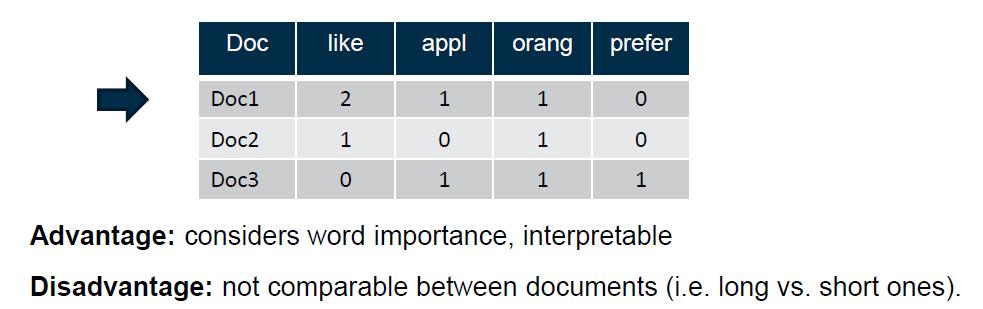




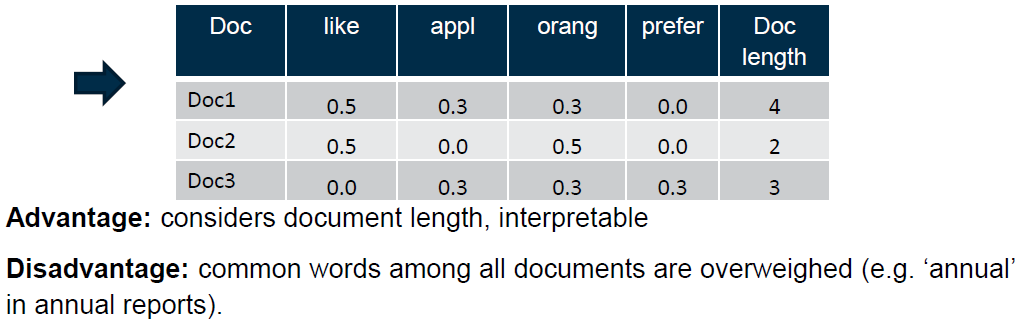
But one-hot-encoding has **disadvantages**:

1. Not possible to compare documents of different length
2. All words are considered equally important
3. Not useful for short texts

# Bag-of-words: Absolute frequency

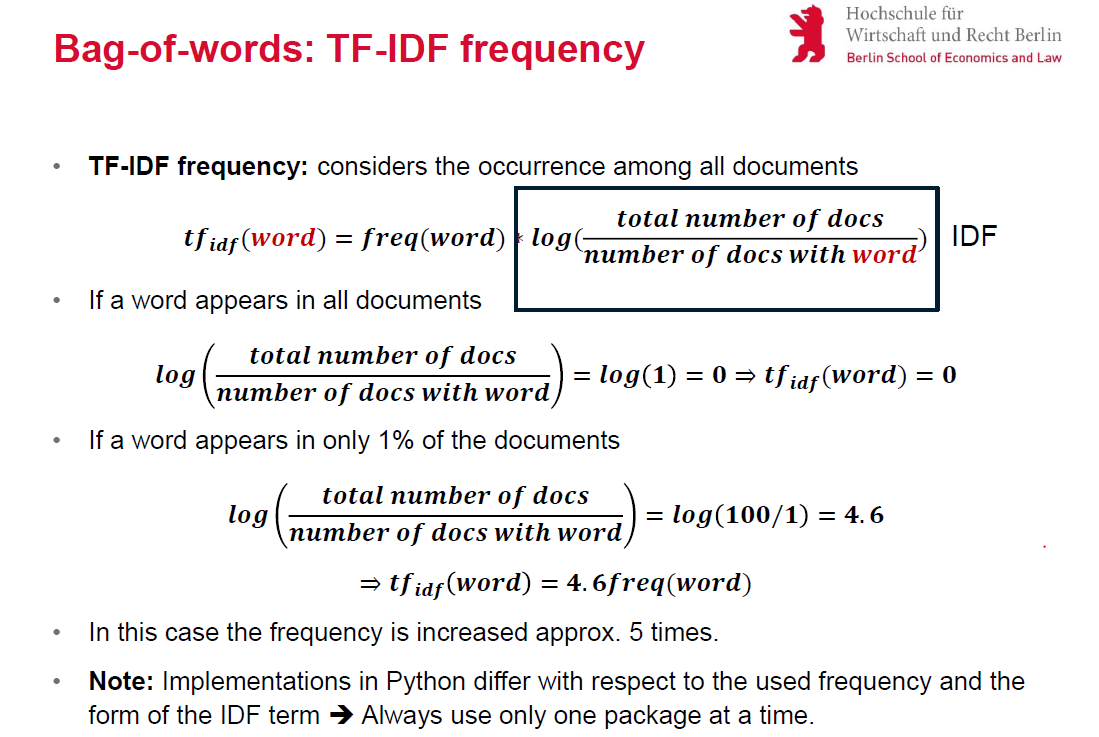


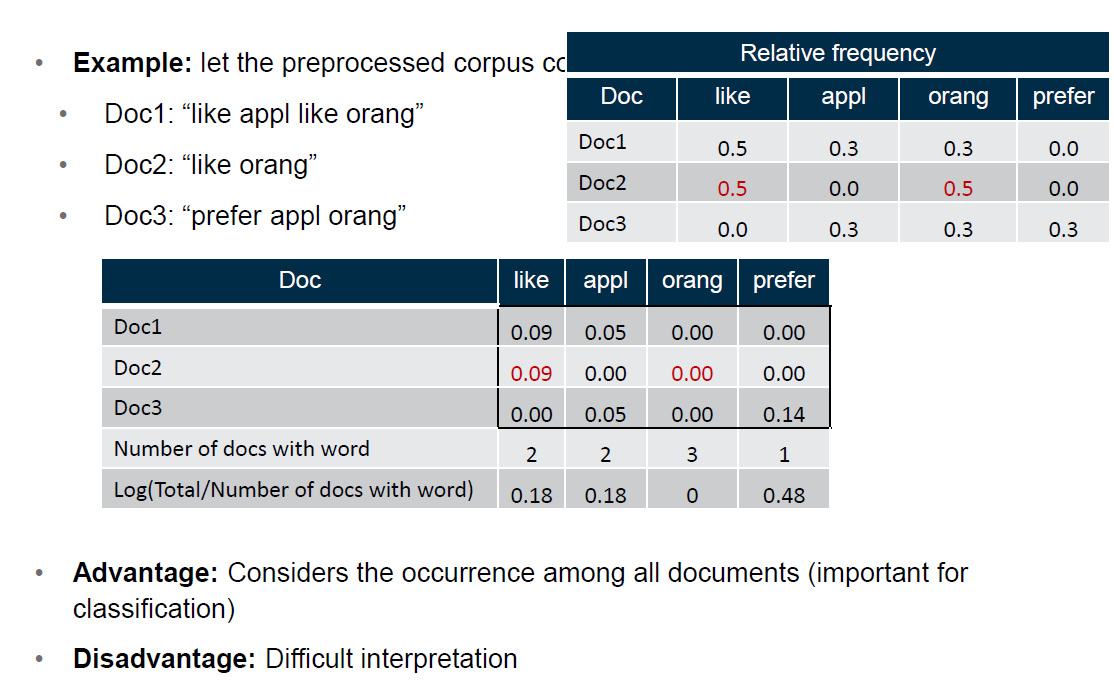
# Bag-of-words: Relative frequency



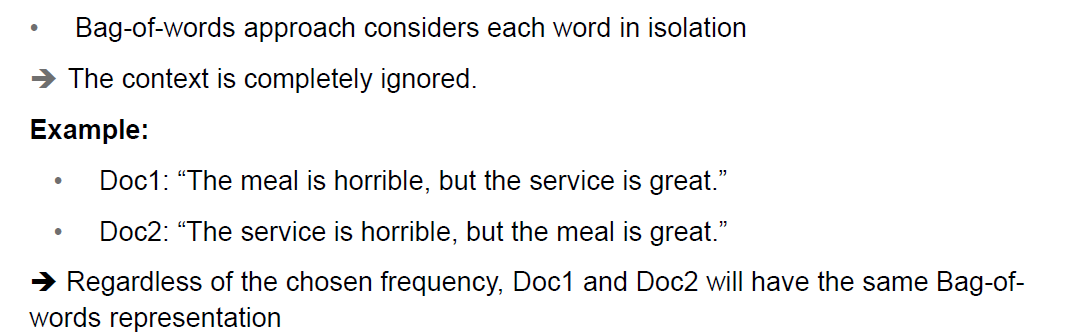
It’s important to not overweight common words:

# Bag-of-words: TF-IDF frequency





# Disadvantages of Bag-of-words:



# N-grams

