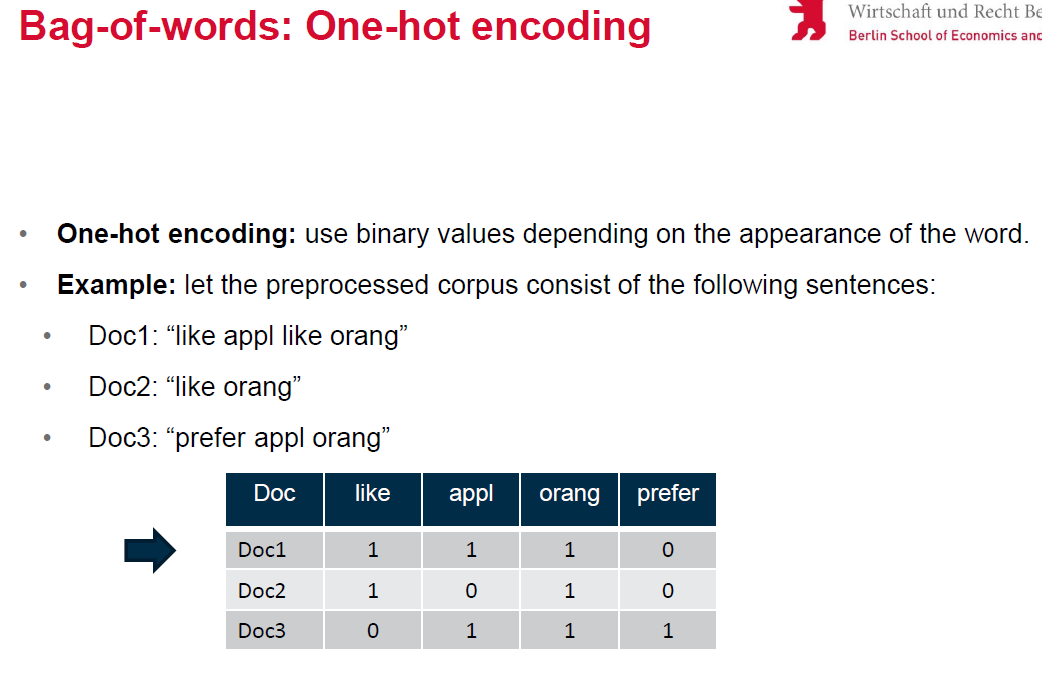
2nd Lesson

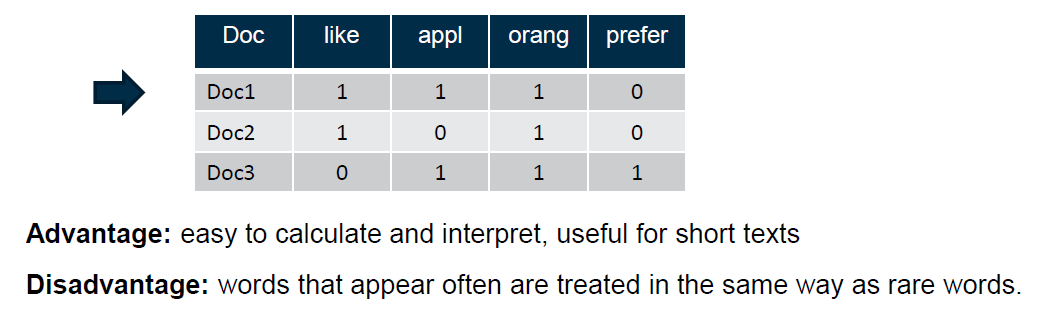
Weakness of Sentiment analysis: negations

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

TF 🡪 Document Frequencies

# Bag-of-words: One-hot encoding

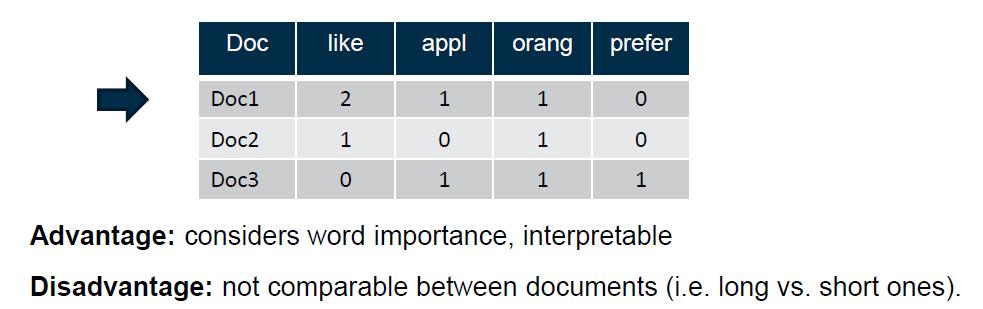




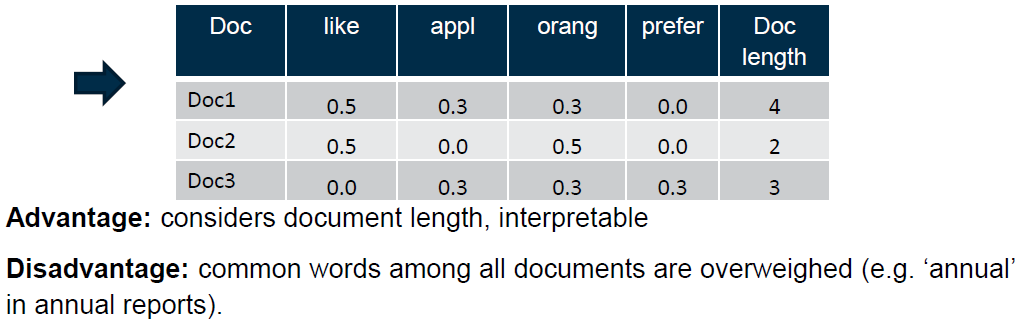
But one-hot-encoding has **disadvantages**:

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

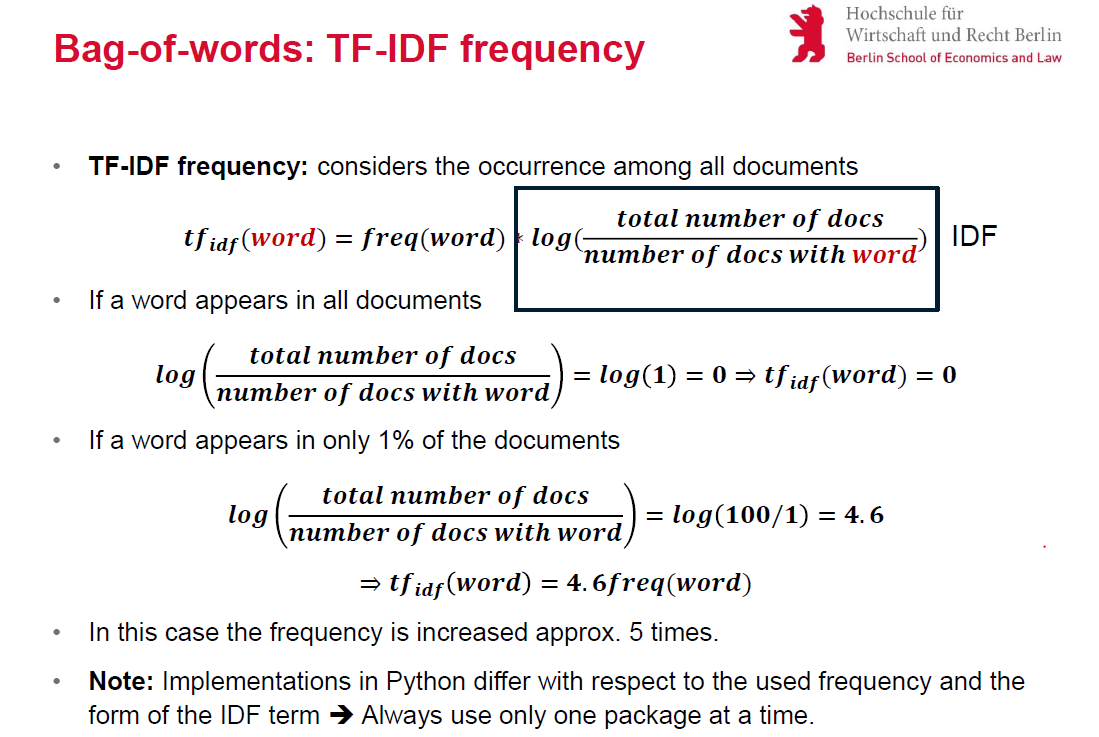


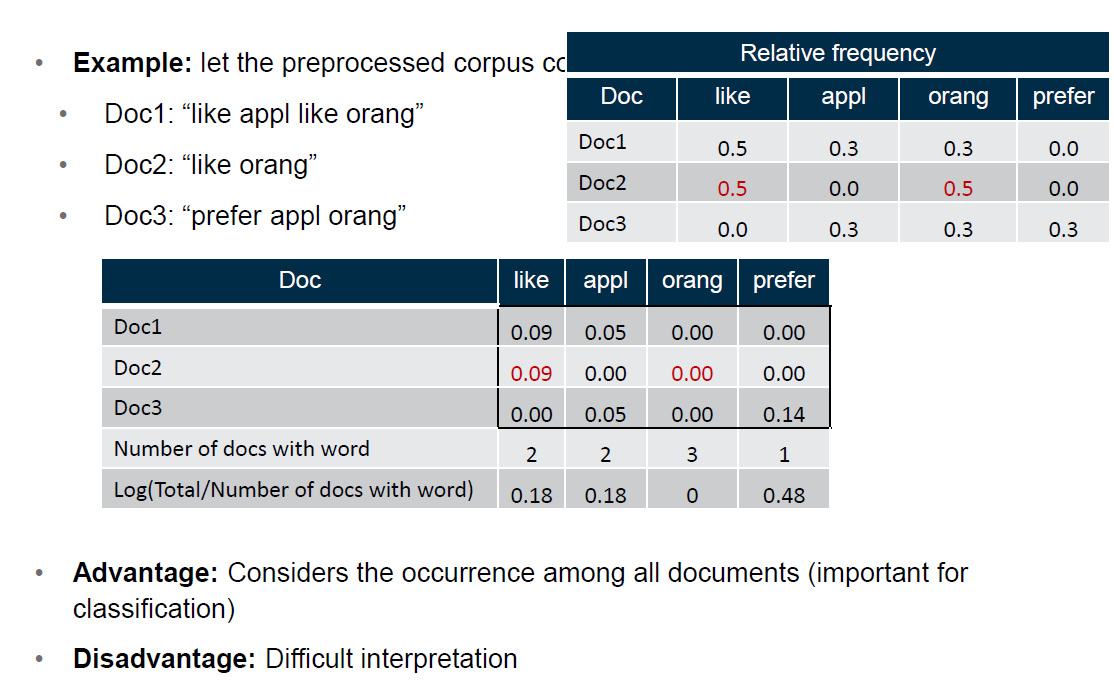
# Bag-of-words: Relative frequency



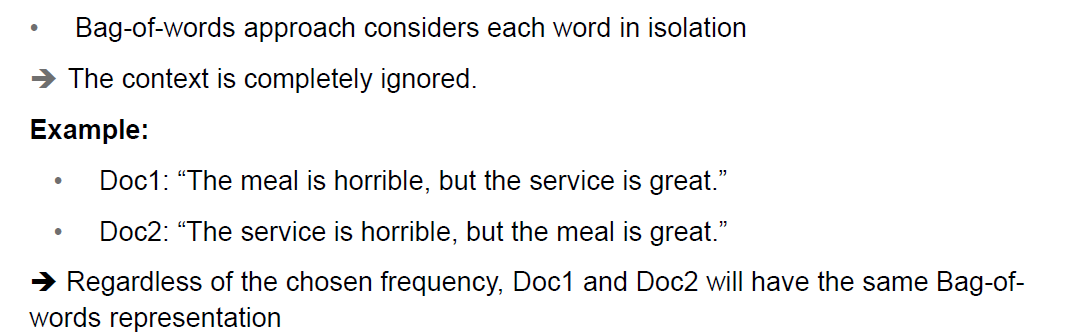
It’s important to not overweight common words:

# Bag-of-words: TF-IDF frequency

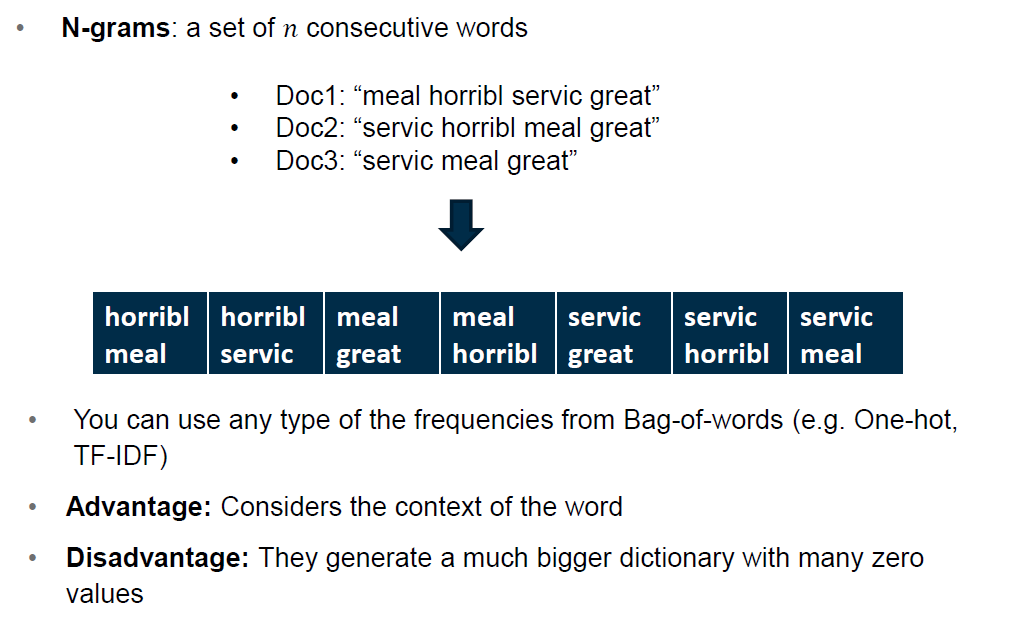




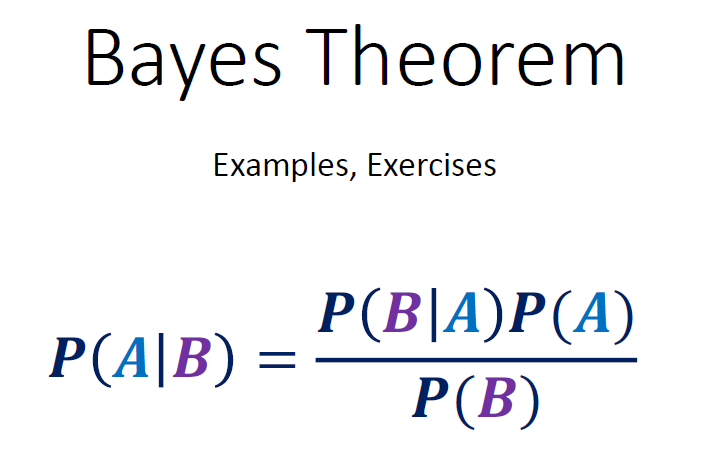
# Disadvantages of Bag-of-words:



# N-grams



3rd Lesson



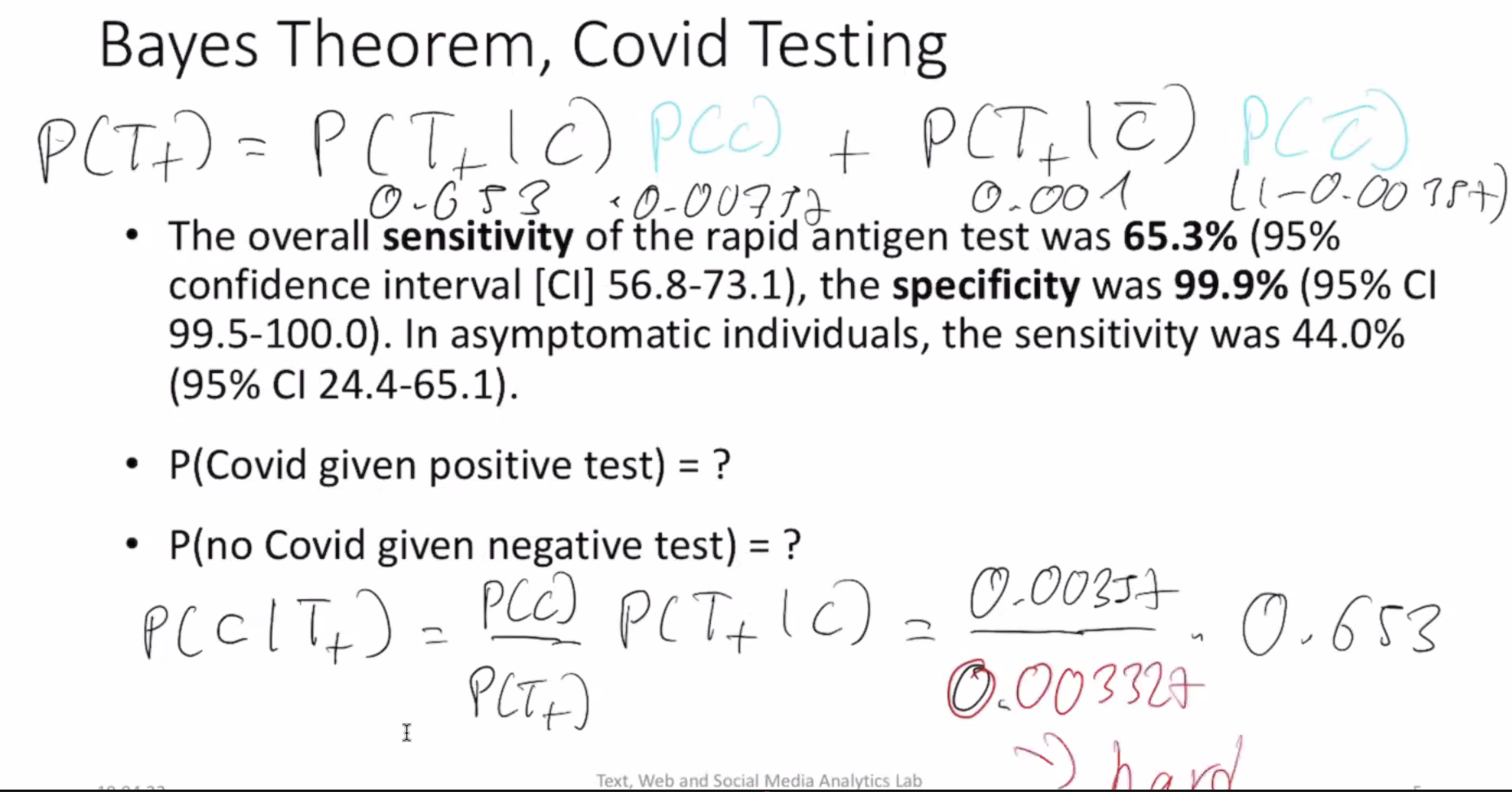
Baseline fallacy:

P(G|A) = 0.001

P(A|G) = P(A) / 0.001 x 0.001 = P(A)

Law of total probability:



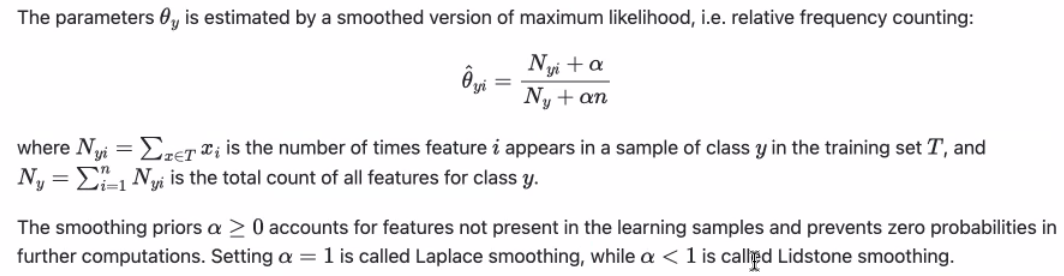


**Naïve bayes**: you never get the true probability out.

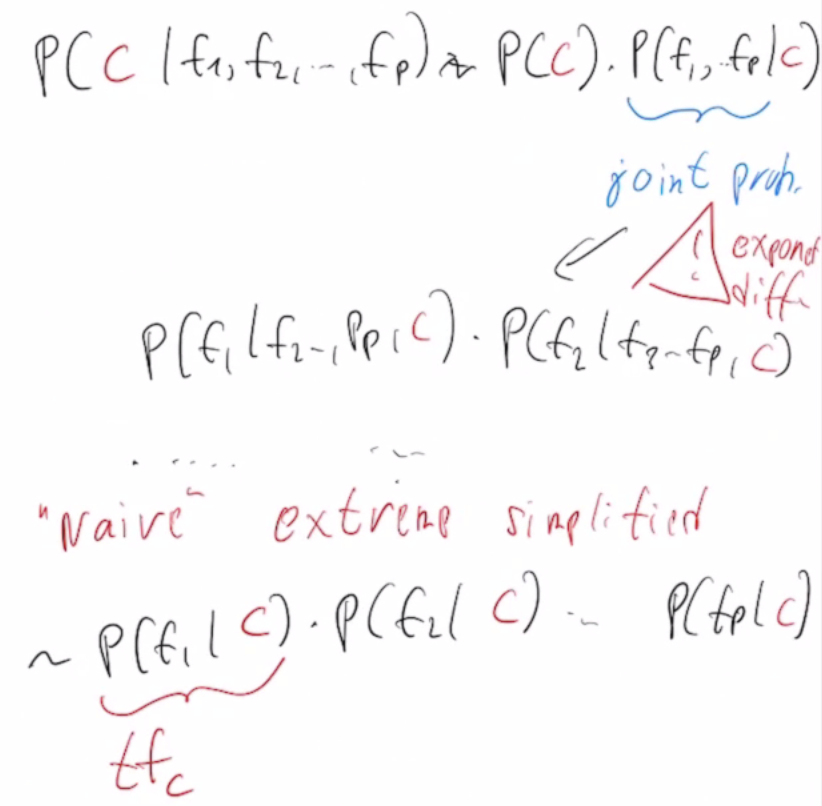
Simplifications:

1. It’s useful to leave the denominator outside because it’s shared between all choices.
2. You have multiple features.

Multinomial Naïve Bayes:



**Low count smoothers:** prior knowledge



**Neural Networks**

Regression = only 1 output node

Multinomial = different output nodes, one for each class

It’s done with the **keras** library

Number of weights 1st step: 10.000 x 16 features + 16 bias

Number of weights 2nd step: 16 x (16+1)