

**For which document classifier the training cost is low and inference is expensive?**

- A. for none
- B. for kNN
- C. for NB
- D. for fasttext

Answer B

For kNN the only cost incurred for training is the computation of document weights, including the idf values. In Naïve Bayes for all words the likelihood of the word to occur in a class has to be computed. Fasttext requires SGD.

**How many among the listed classifiers can be used to derive probability estimate of the class label?**

1NN, kNN, Rocchio, NB, fasttext

- A. 1
- B. 2
- C. 3
- D. 4

Answer C

1NN cannot be used to compute a probability estimate, since it provides a binary decision of which is the closest class label. The same is true for the Rocchio method.

kNN can be used to estimate the probabilities for class labels if  $k$  is sufficiently large. NB and fasttext by definition provide probability estimates.

## Question

The number of parameters of the fasttext classifier and the simple self-attention classifier

1. Are the same
2. Fasttext has more
3. Self-attention has more

## Answer 1

The simple self-attention based classifier applied before the averaging of the word vectors the self-attention mechanism. This mechanism uses only the word vectors as weights and therefore does not introduce new parameters into the model.