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

KNN training cost is relatively low, but in the inference, for every sample we have to look into whole training data to look for k nearest neighbor, and therefore the inference is expensive. Therefore, the second answer is correct.

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

fastText and NB can be used to derive probabilities. Also for KNN, we can use it for large k values to estimate probabilities. Rocchio and 1NN can't be used for probability estimation, and therefore the third answer is correct.

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

In both models the only parameters are the parameters used to store (sub)word embeddings, and therefore the number of parameters are the same. So, the correct answer is the first answer.