

Nearest Neighbour LMs

Advanced NLP: Summer 2023

Anoop Sarkar

GENERALIZATION THROUGH MEMORIZATION: NEAREST NEIGHBOR LANGUAGE MODELS

Urvashi Khandelwal^{†,*}, Omer Levy[‡], Dan Jurafsky[†], Luke Zettlemoyer[‡] & Mike Lewis[‡]

[†]Stanford University

[‡]Facebook AI Research

`{urvashik, jurafsky}@stanford.edu`

`{omerlevy, lsz, mikelewis}@fb.com`

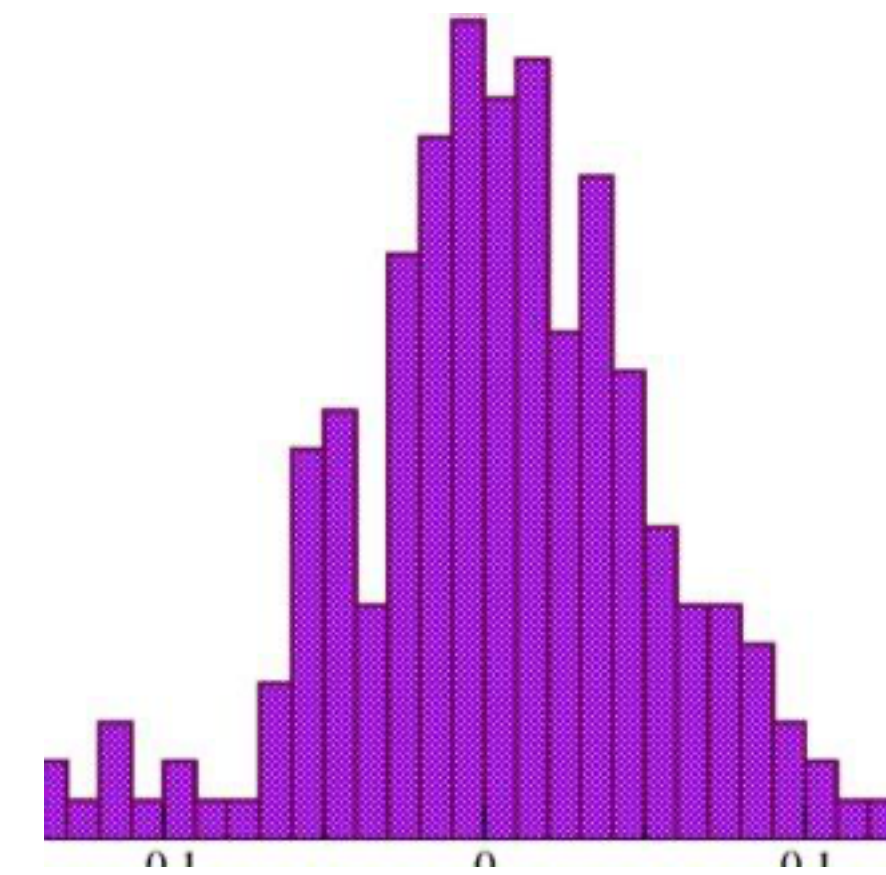
<https://openreview.net/forum?id=HkIBjCEKvH>

Learning representations is easier than prediction

Dickens is the author of

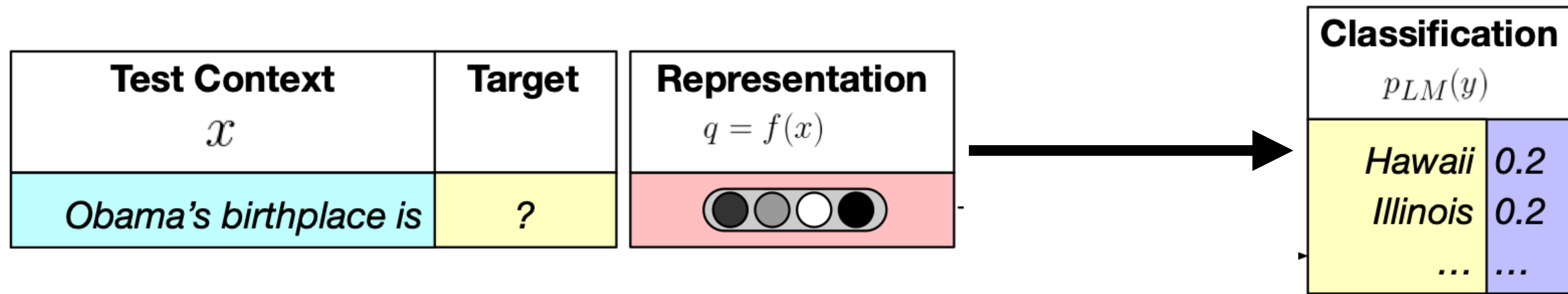
Dickens wrote

?



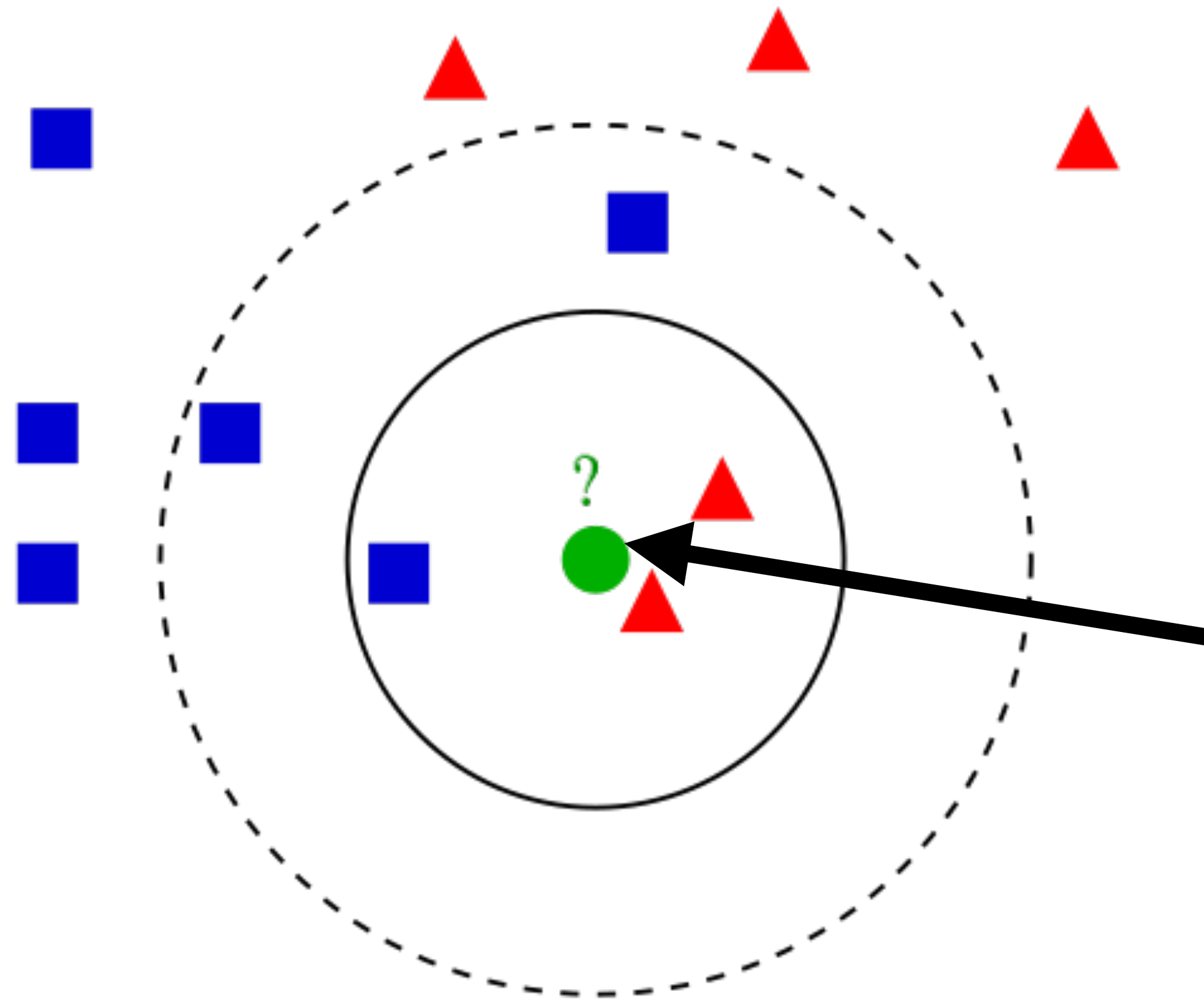
Even if you cannot predict the next token, you can predict that the distribution is identical over the vocabulary.

Standard LM prediction



Nearest Neighbour

k neighbours in vector space



The query vector is compared to other data vectors in the same vector space.

Choose the class that has the most representatives inside the search perimeter.

kNN LM prediction

