## Nearest Neighbour LMs

**Advanced NLP: Summer 2023** 

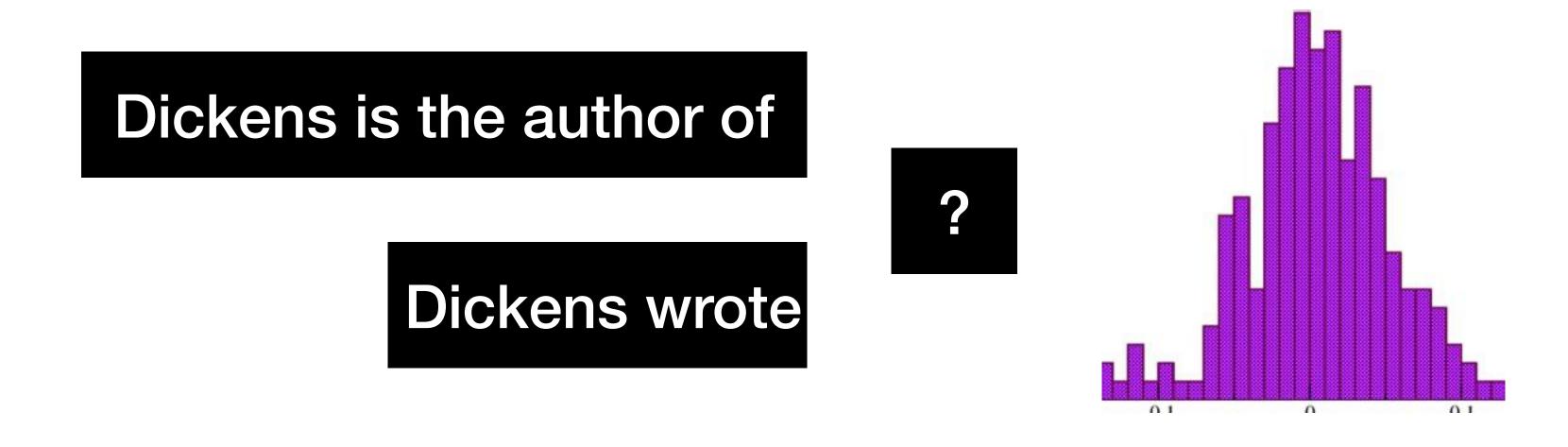
# GENERALIZATION THROUGH MEMORIZATION: NEAREST NEIGHBOR LANGUAGE MODELS

```
Urvashi Khandelwal<sup>†</sup>, Omer Levy<sup>‡</sup>, Dan Jurafsky<sup>†</sup>, Luke Zettlemoyer<sup>‡</sup> & Mike Lewis<sup>‡</sup> Stanford University
```

<sup>‡</sup>Facebook AI Research {urvashik, jurafsky}@stanford.edu {omerlevy, lsz, mikelewis}@fb.com

https://openreview.net/forum?id=HklBjCEKvH

#### Learning representations is easier than prediction



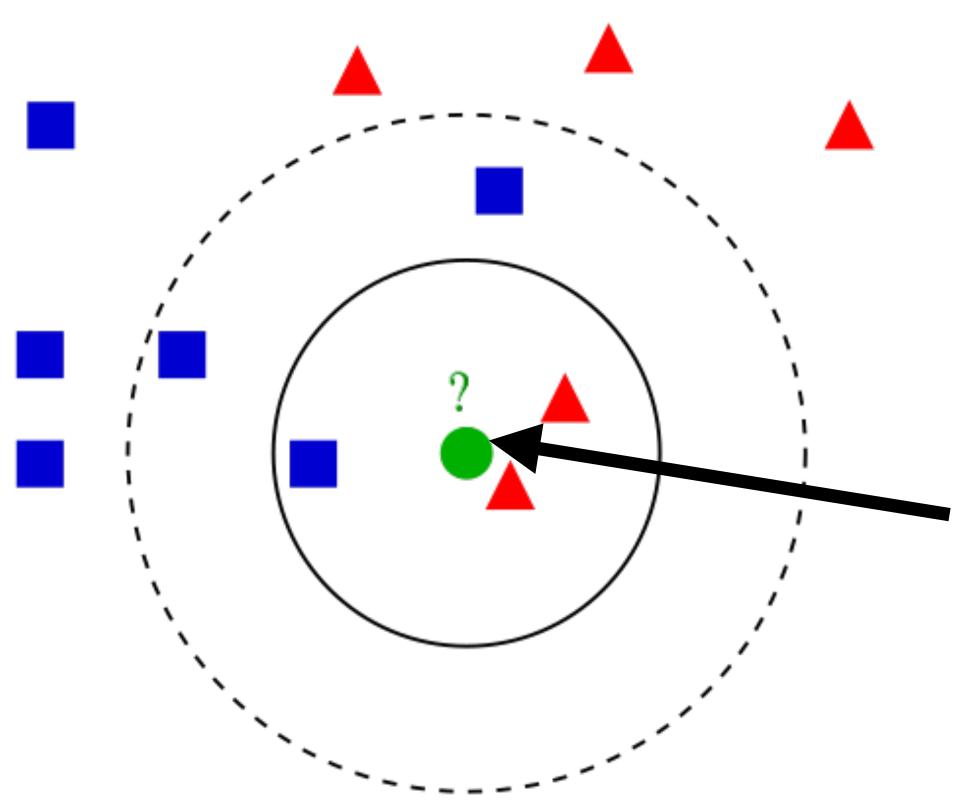
Even if you cannot predict the next token, you can predict that the distribution is identical over the vocabulary.

#### Standard LM prediction

			_	Classification		
Test Context	Target	Representation		$p_{LM}(y)$		
$\boldsymbol{x}$		q = f(x)		Hawaii	0.2	
Obama's birthplace is	?		-	Illinois		-
			•	• • • •		

#### Nearest Neighbour

k neighbours in vector space

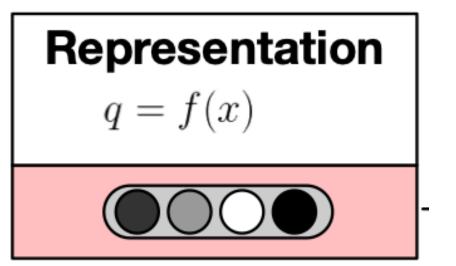


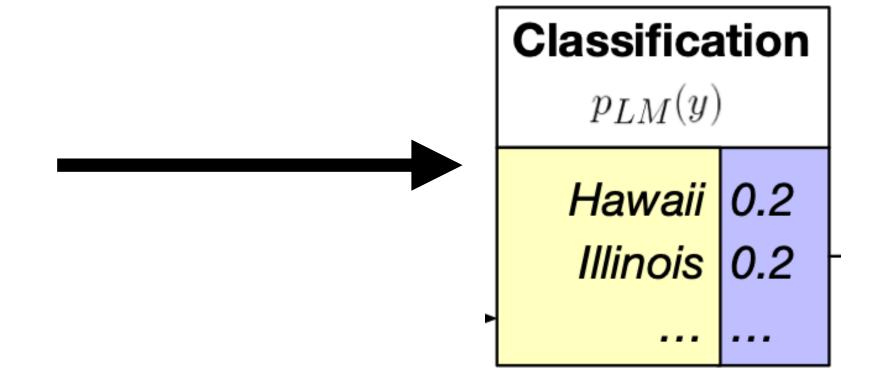
The query vector is compared to other data vectors in the same vector space.

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

### kNN LM prediction

Test Context $x$	Target	
Obama's birthplace is	?	





Training Contexts $c_i$	Targets $v_i$	Re
Obama was senator for Barack is married to Obama was born in	Michelle Hawaii 	
Obama is a native of	Hawaii	

