Supervised machine learning for text classification

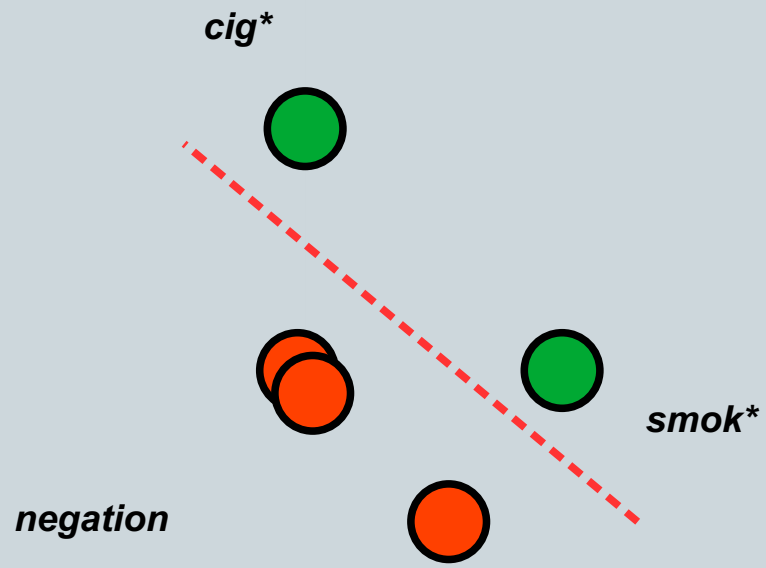
NLP in one day

KING'S College LONDON

NIHR | Maudsley Biomedical Research Centre

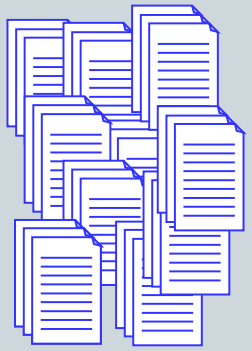
Learning from labelled examples

Example sentence	Label	Features		
		smok*	cig*	negation
He smokes	T	1	0	0
Suffers from anhedonia	F	0	0	0
She does not smoke	F	1	0	1
20 cigarettes a day	T	0	1	0
Blood pressure 70/120	F	0	0	0



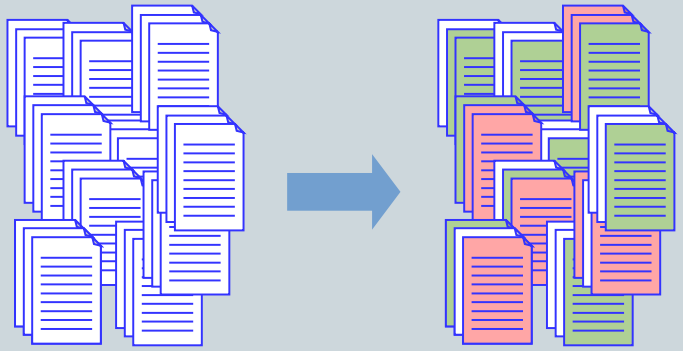
- Compute some line or plane separating our positive and negative examples
- Use this plane to determine the class (label) of previously unseen examples
- This is *supervised classification*

Supervised classification



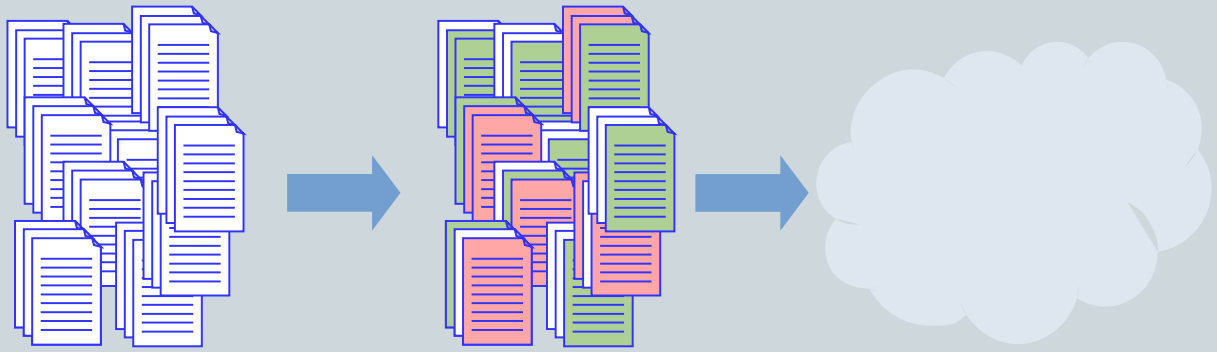
- Take a set of example texts.
- They might be sentences, whole documents, single words, or some other portion of text.
- This is our training corpus.

Supervised classification



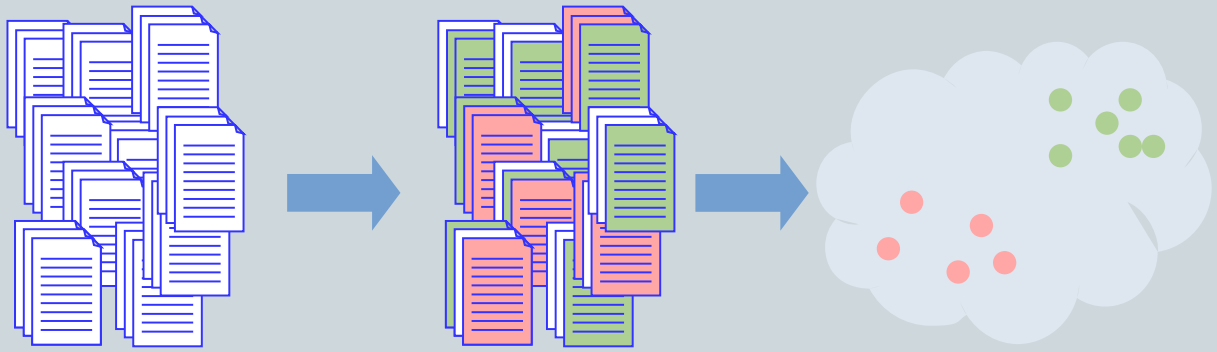
- Label each example, with the classes in our problem.
- Labelling will often be done by human.
- We might be lucky enough to have some existing labelled data, e.g. radiology reports with a code for tumour class attached..

Supervised classification



- Select features to represent our texts.
- These might be the presence of words, POS tags, distances between words, word sequences (ngrams), presence of word groups, sentence lengths, etc.
- We may use numeric representations of words as features, computed in a separate step. In the state of the art, these are referred to as embeddings.

Supervised classification



- Represent the texts in this feature space.

Supervised classification

- Compute some separator between classes.
- This will involve measures of distance between points.
- It might also involve methods for projecting multiple dimensions into different spaces in which they are separable (kernels).

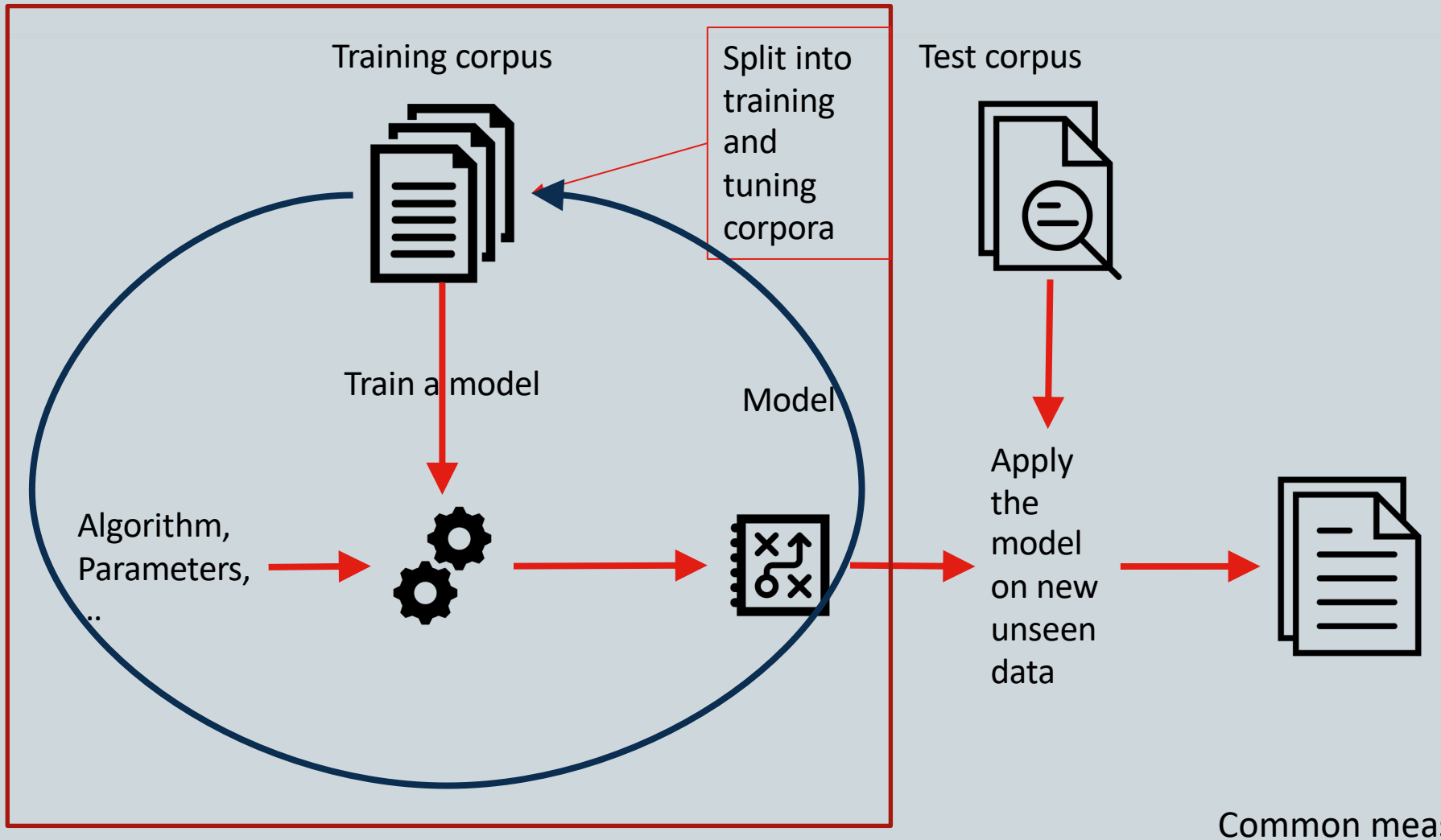
Supervised classification

- Commonly used classification algorithms in NLP:
  - SVM (very popular)
  - CRF
  - Naive Bayes
  - KNN
  - Random Forest
  - State of the art: neural nets, e.g. CNNs, LSTMs

Supervised classification

- Classify / label new, previously unseen examples by representing them in the same feature space.

How good is the model?



Common measures:

- Precision, P == positive predictive value
- Recall, R == sensitivity
- F1, the harmonic mean of P and R

Thank you

angus.roberts@kcl.ac.uk
https://www.kcl.ac.uk/people/angus-roberts