Why Are Linear Text Classifiers Popular?

School of Information Studies
Syracuse University

K-Nearest Neighbor (k-NN)

Training process:

- Add in all training examples.

Classification process:

- Given a new example x, compare the similarity between x and all training examples, and choose the majority-voted category label in the k nearest training examples.

K-Nearest Neighbor

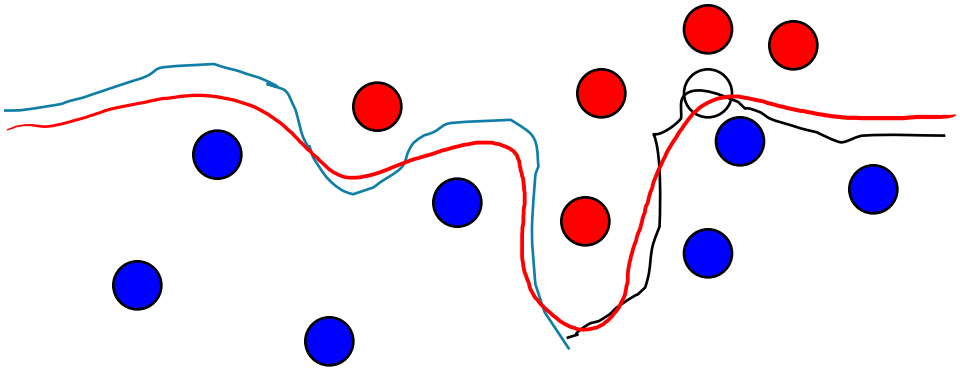
Lazy learner: no learning, just predicting

Instance-based learning

Advantages of k-NN

No assumptions: non-parametric method

When the target function to be learned is very complex



Disadvantages of k-NN

Sensitive to noisy training data

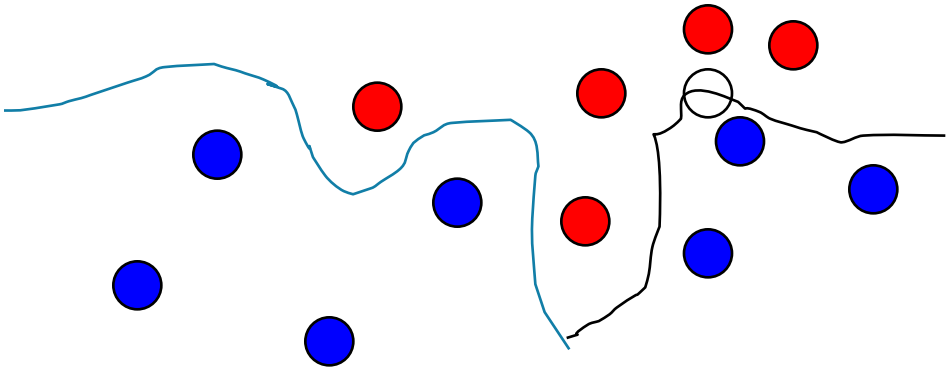
- All attributes participate in classification.
- If only a few attributes are relevant to prediction, the participation of those irrelevant attributes would harm the prediction performance.

High computational cost

- Algorithms like Naïve Bayes would create a model on the training data, and then use the model only to predict new data, which is fast.
- For k-NN, since there is no training step, nearly all computation takes place in the prediction step. If there are many examples and many attributes in the data set, computational cost is high.

The Shape of Decision Function

No regular shape: k-NN

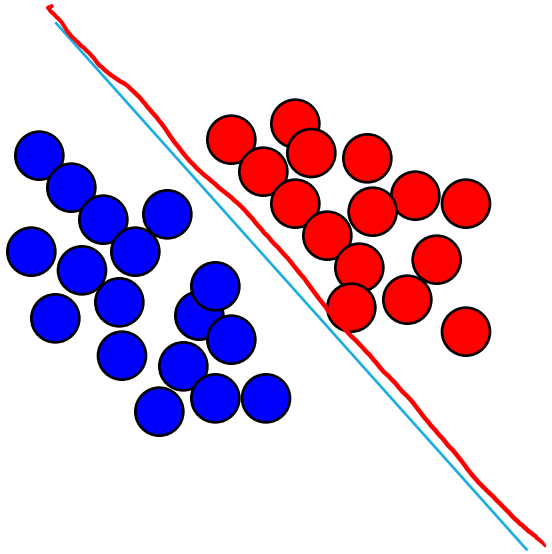


The Shape of Decision Function

Linear: Naïve Bayes, SVM

How many parameters to determine a line in 2D space?

- Y = ax+b
- Weight
- Intercept



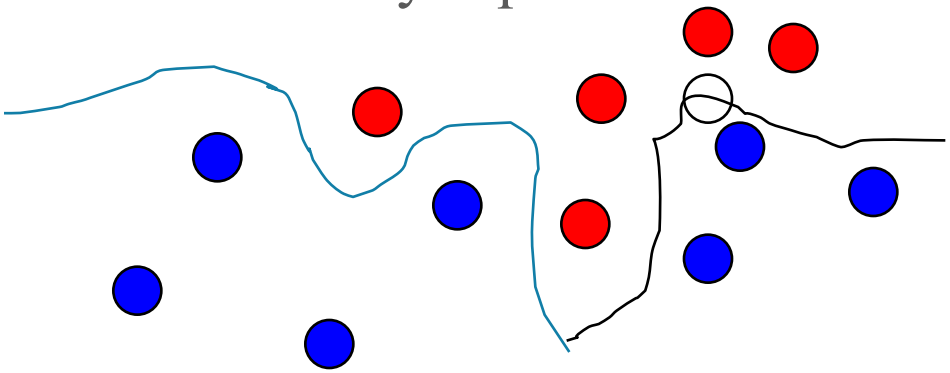
Why Are Linear Classifiers So Popular for Text Classification?

Most text classification problems are linearly separable.

- Large number of features
  - Example: 16K word features in the movie review data
- Usually, a linear boundary can be found to separate the data

The Shape of Decision Function

Some data are not linearly separable



Use non-parametric method like k-NN