

Can you classify between **muffin** and **dog** ?



# When do we use KNN?

We can use K Nearest Neighbor when



Dataset is small

Because KNN is a “lazy learner”

Dataset is noise free

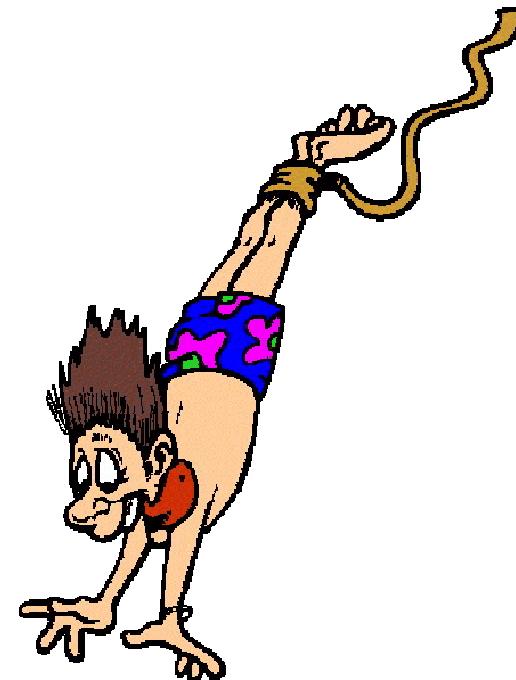
Weight	Height	Class
51	182	Underweight
62	165	One - forty
69	176	Hello
64	173	23
65	172	Normal



Data is labelled

Dog

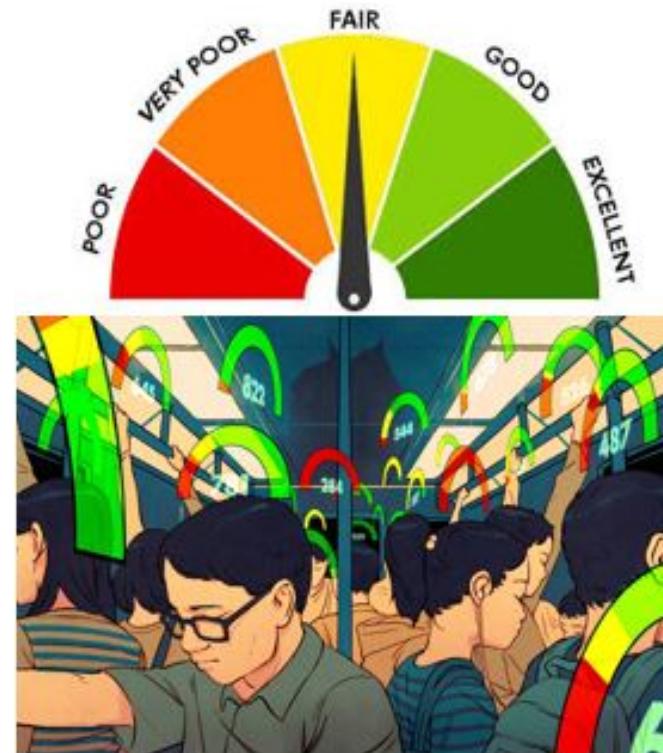
Noise



# Applications

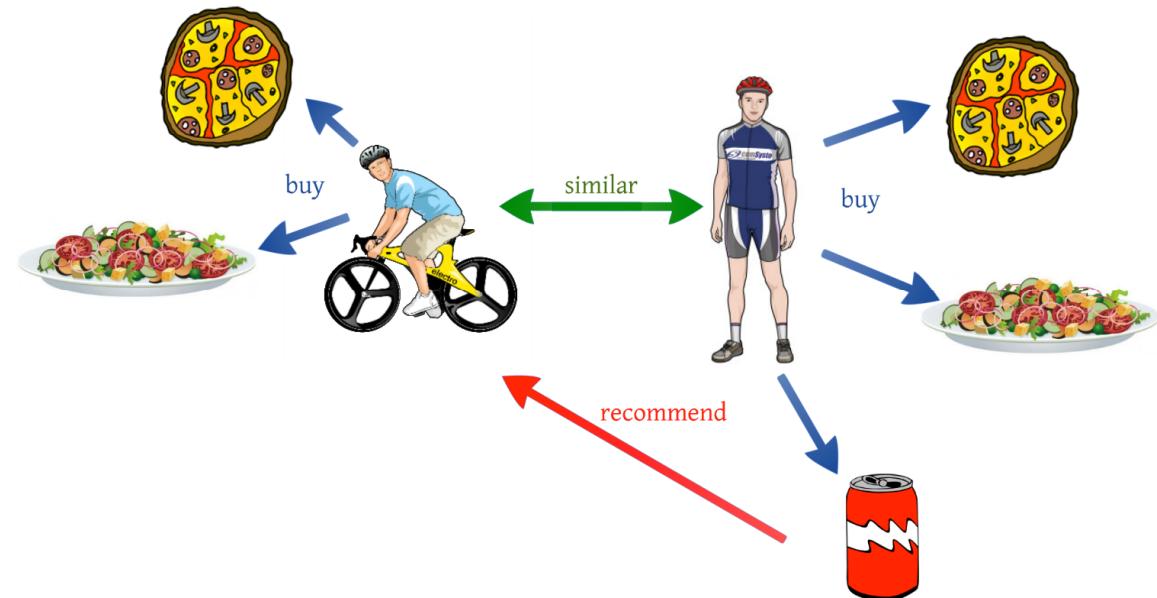
# Credit Ratings

- Collecting financial characteristics vs. comparing people with **similar financial features** to a database.
- By the very nature of a credit rating, people who have **similar financial details** would be given similar credit ratings.



# Recommender system

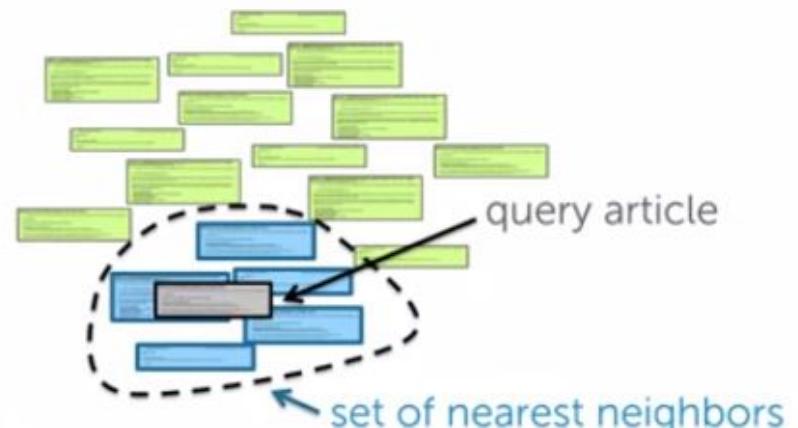
- If you know a user likes a particular item, then you can **recommend similar items** for them.



# Document Retrieval

- We compute **distances from query article** to all other articles.
- Then we search for the articles with **smallest distance** to the query article.
- They are called **nearest neighbors**.

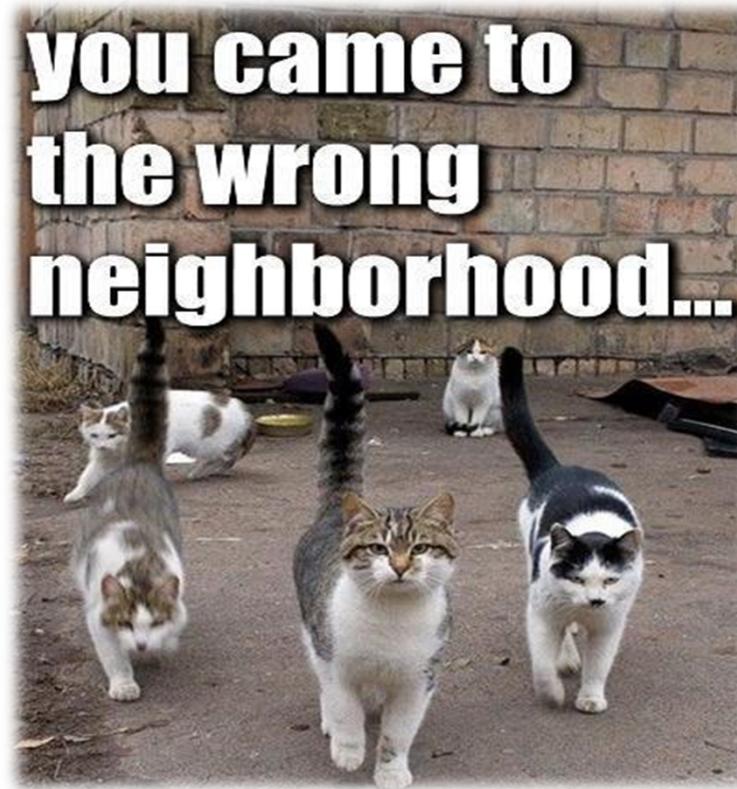
Space of all articles,  
organized by similarity of text



# Advantages of K-nearest neighbors algorithm

- It is **simple** to **implement**.
- It **executes quickly** for small training data sets.
- Performance **asymptotically approaches** the performance of the Bayes Classifier.
- Don't need any prior knowledge about the **structure of data** in the training set.
- **No retraining** is required if the new training pattern is added to the existing training set.

# Limitations



# Limitations of K-nearest neighbors algorithm

- When the training set is large, it may take a lot of space.
- For every test data, the distance should be computed between test data and all the training data. Thus a lot of time may be needed for the testing.