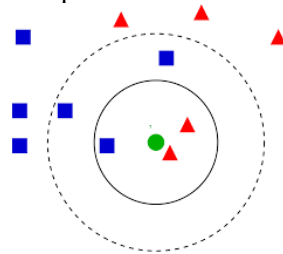


9.1 What's K-Nearest Neighbor?	1
9.2 Algorithm	1
9.3 Pros and Cons	1
9.4 KNN in Practice	1

9.1 What's K-Nearest Neighbor?

K-Nearest Neighbor (KNN) is supervised algorithm in machine learning. The core idea is that if most of the **k most adjacent samples of a sample** in the feature space belongs to a certain category, then the sample also belongs to that category. In other words, each sample can be represented by its closest K neighbors.

As the figure shown below, which category should the green circle be assigned? A red triangle or a blue square? If K is equal to 3, the green circle will be assigned to the red triangle because the proportion of the red triangle is 2/3; if K is equal to 5, the green circle will be assigned to the blue square because the proportion of the blue square is 3/5.



9.2 Algorithm

- 1) calculate the distance between test data and each training data.
- 2) sort according to the increasing relation of distance.
- 3) select K points with the smallest distance.
- 4) determine the occurrence frequency of the category of the first K points.
- 5) return the category with the highest frequency in the first K points as the prediction classification of the test data.

9.3 Pros and Cons

Pros:

- Simple and easy to implement.
- There's no need to build a model, adjust several parameters, or make additional assumptions.
- Versatile: It can be used for classification, regression.

Cons:

- The algorithm gets significantly slower as the increase of the number of examples or predictors or independent variables.

9.4 KNN in Practice

At scale, this would look like recommending products on Amazon, articles on Medium, movies on Netflix, or videos on YouTube. Although, we can be certain they all use more efficient means of making recommendations due to the enormous volume of data they process. However, we could replicate one of these recommender systems on a smaller scale using what we have learned here. Let us build the core of a movies recommender system.