k-NN

k-nearest neighbors algorithm


KNN BE LIKE
"Show me your friends, and I'll tell you who you are."

Requirements

- Some notion of distance
- An assumption that points that are close to one another are similar

Importance of *k*

AKA hyperparameter

The whole algorithm is based on the *k* value. Even small changes to k may result in big changes.

Easiest way to choose *k* is based on accuracy of prediction in training.

Steps

KNN has three basic steps

1. Calculate the distance.
2. Find the k nearest neighbours.
3. Vote for classes

Distance measurement

A few of possible techniques

1. Euclidean distance
2. Hamming distance
3. Mahalanobis distance

The Curse of Dimensionality

k-nearest neighbors runs into trouble in higher dimensions thanks to the "curse of dimensionality," which boils down to the fact that high-dimensional spaces are vast. Points in high-dimensional spaces tend not to be close to one another at all.

Code: repository

Resources

---

1. https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm
2. https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761
3. https://www.oreilly.com/library/view/data-science-from/9781491901410/