Aprendizagem Computacional
Machine Learning

K-Nearest Neighbours

Departamento de Engenharia Informática,
Universidade de Coimbra

Catarina Silva, 2024

Contents

- Introduction
- Algorithm
- Applications

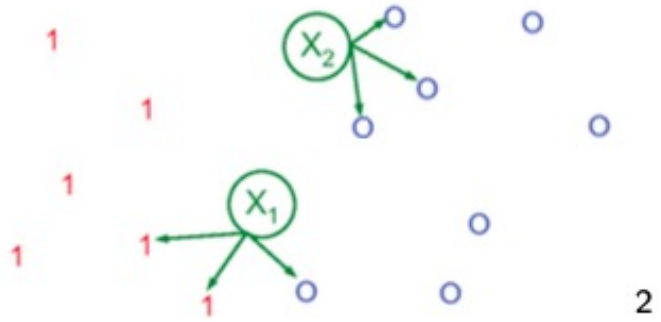
KNN – K Nearest Neighbors

- Lazy learning (no model)
- Instance-based learning

KNN Definition

Training: store all training examples (perfect memory)

Test: predict value/class of an unseen (test) instance based on closeness to stored training examples, relative to some distance (similarity) measure



Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*

KNN Algorithm (1)

- For K=1,
  - predict the same value/class as the **nearest** instance in the training set.

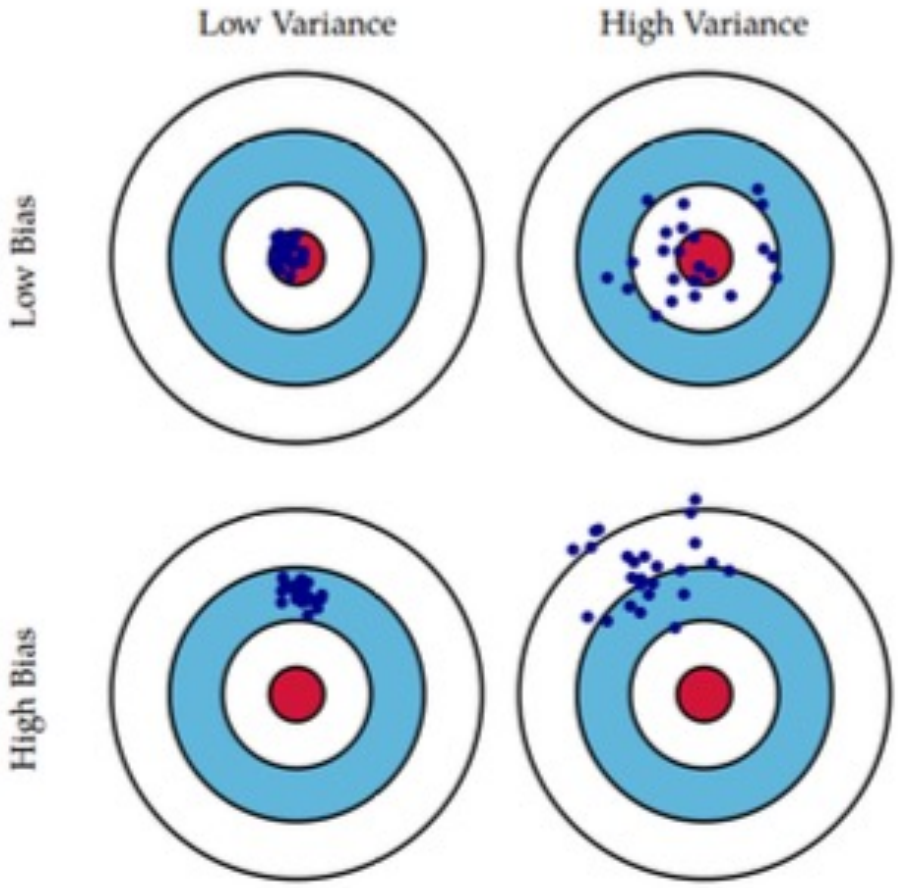
- For K>1,
  - find the K closest training examples, and either
  - predict class by **majority vote** (in classification).
  - predict value by **average weighted inverse distance** (in regression).

KNN Algorithm (2)

- Different distance measures can be used, e.g. Euclidian distance (possibly need to scale/normalize/standartize)
- To avoid ties, use an odd value of K
- Smaller K:
  - predictions have higher **variance** (less stable)
  - good at capturing fine-grained patterns
  - may overfit
- Larger K:
  - predictions have higher **bias** (less true)
  - stable predictions by averaging over lots of examples
  - may underfit, i.e. fail to capture important regularities

- Rule of thumb: k < sqrt(n), n is the number of training examples

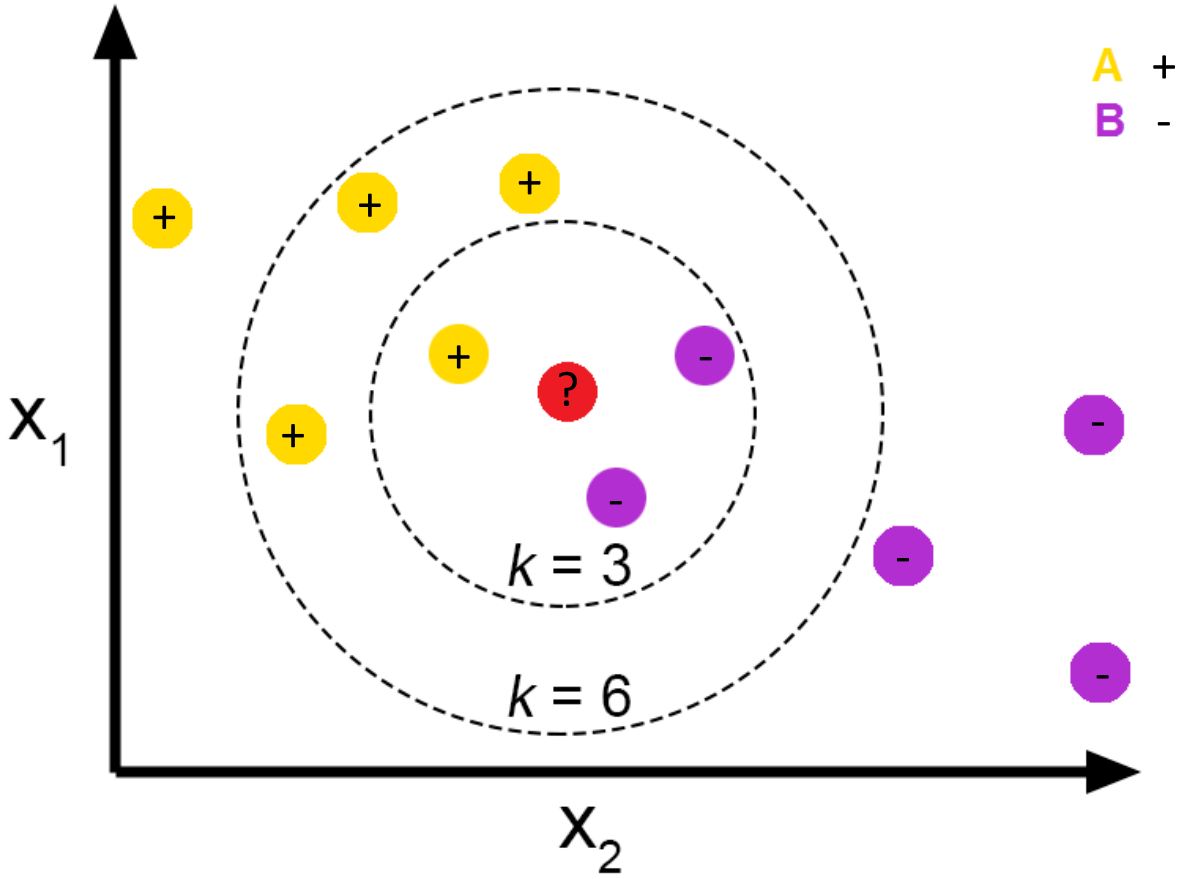
The Bias-variance Tradeoff



- **Biased** models occur when the predictive distribution of the models differs greatly from the target distribution.
- High **variance** models occur when the models have greatly different test predictions (across different training sets from the same target distribution).

Medium.com

Example

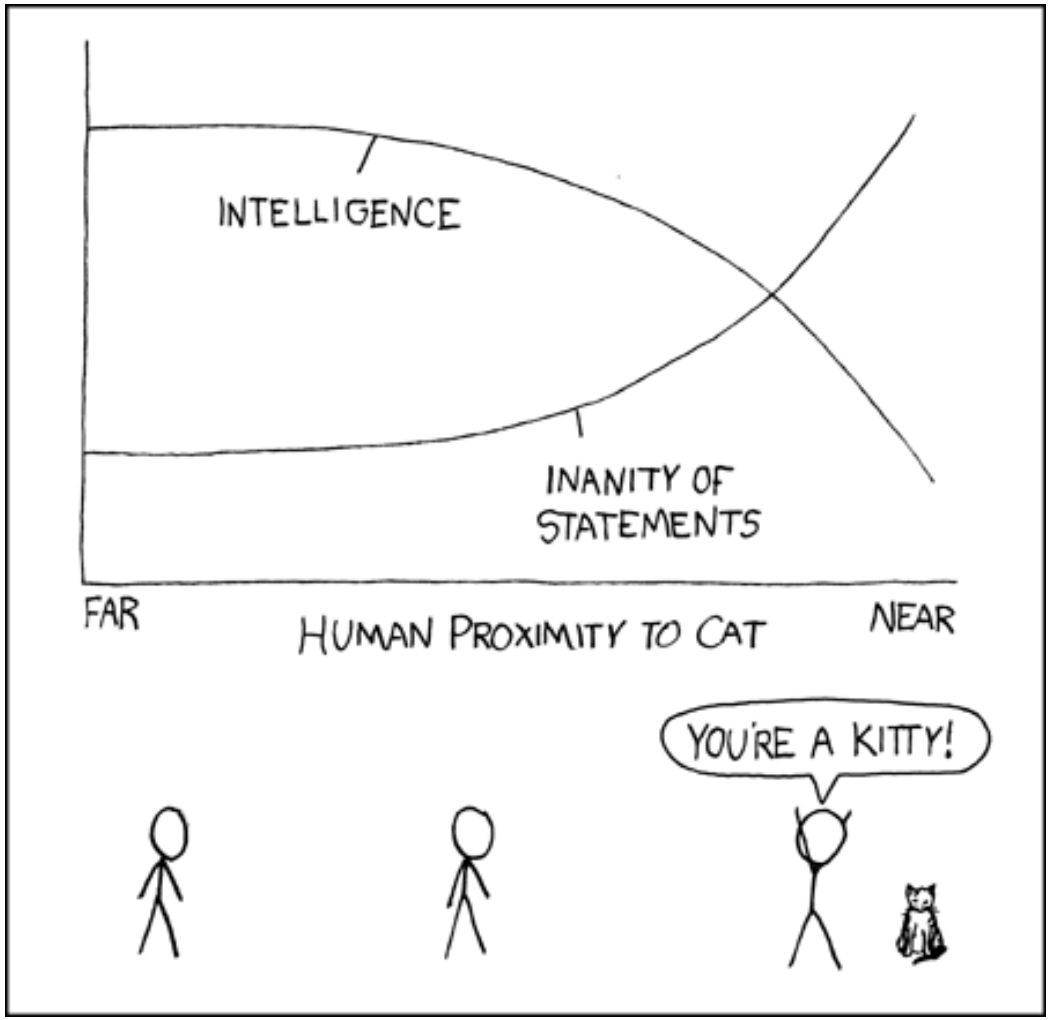


Medium.com

KNN – final considerations

- Simple algorithm, no model, does all its work at test time — in a sense, no learning!
- Can control the complexity by varying k
- Suffers from the Curse of Dimensionality
- Local approximations, interesting for complex goals
- Dynamic by nature, new examples incorporated
- Interpretable

- Not compact (lazy)
- Huge memory and computation requirements



https://xkcd.com/231/