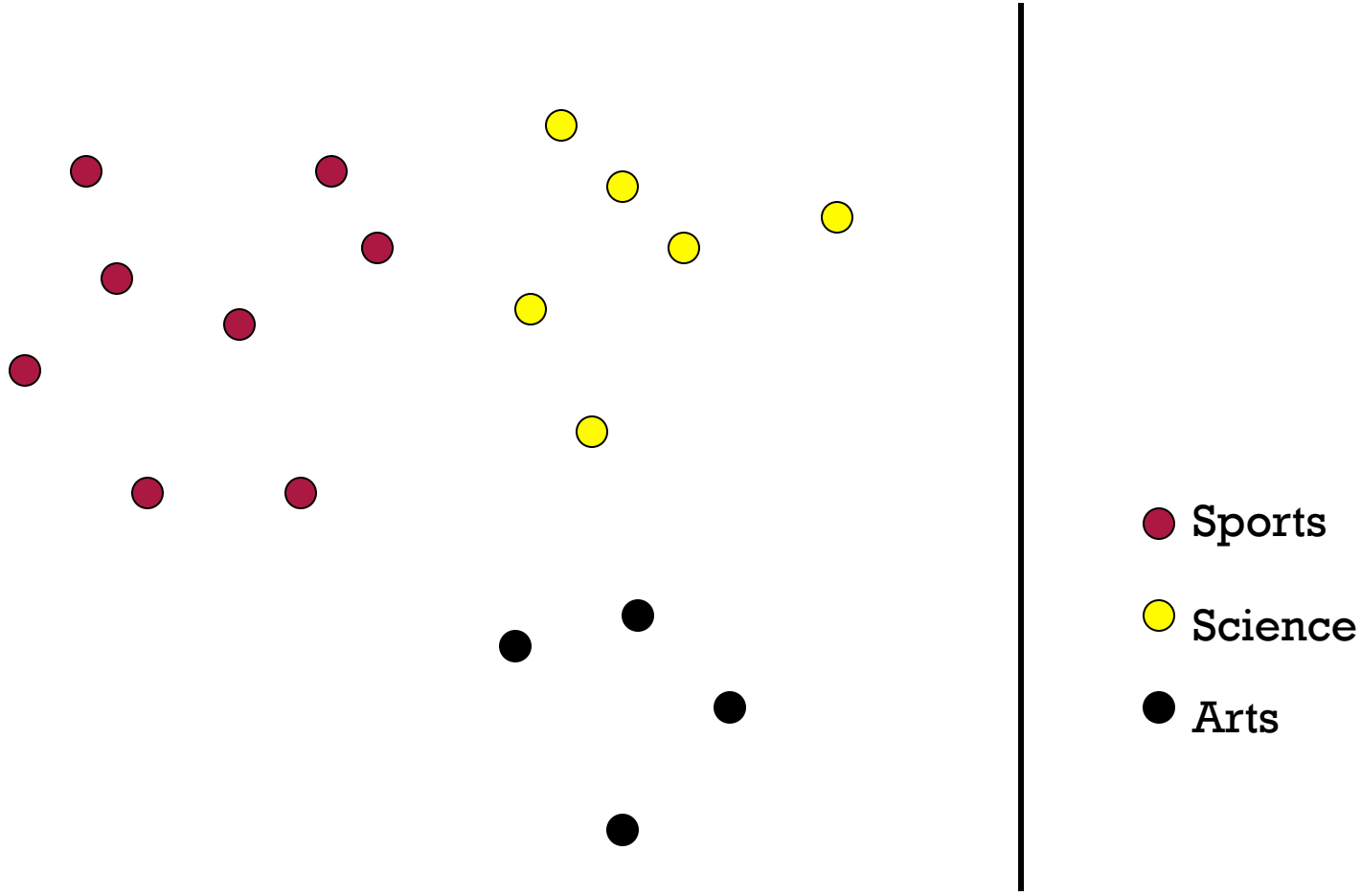
k-NN (k-Nearest Neighbor) Classifier

Aarti Singh

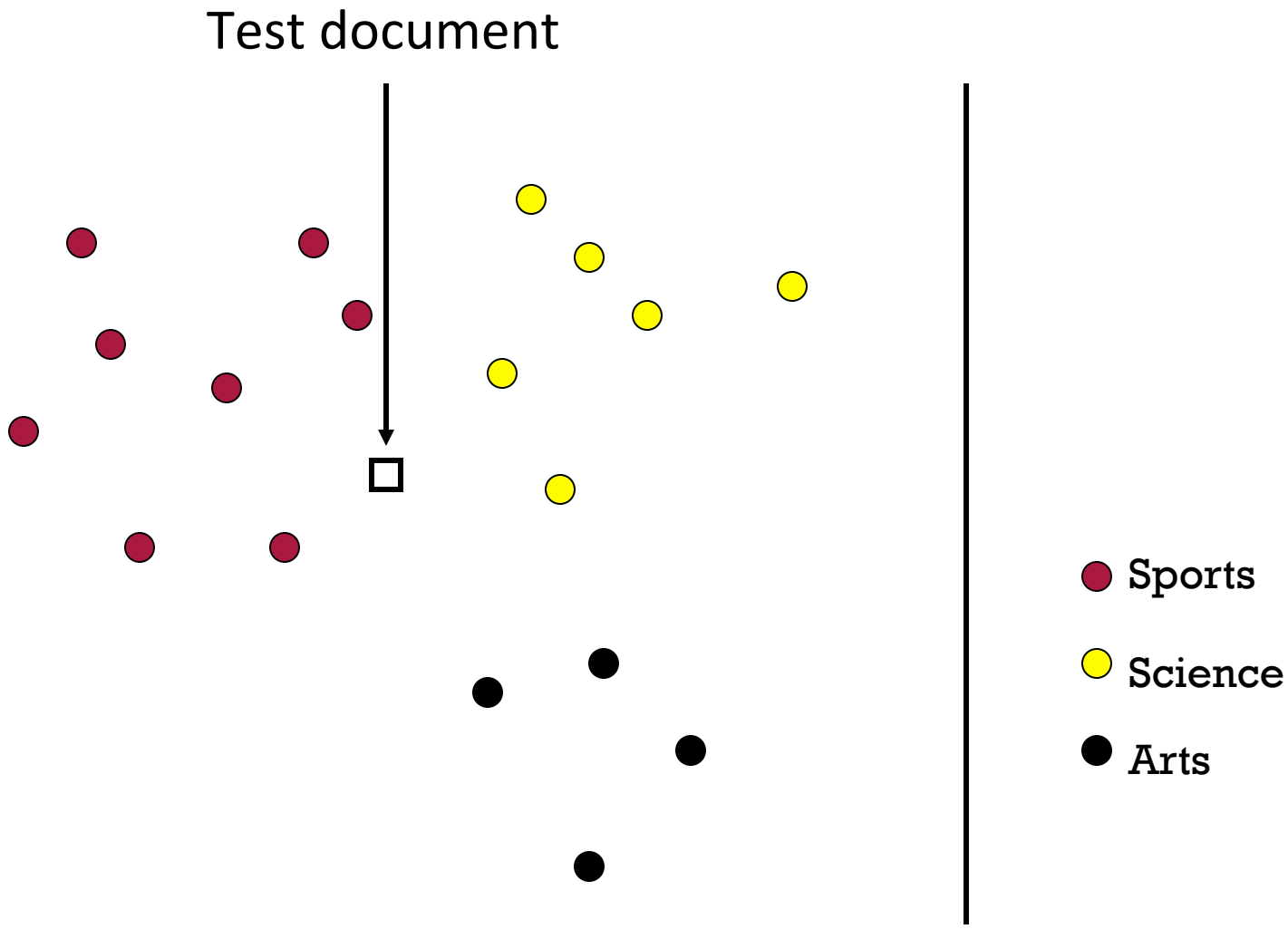
Co-instructor: Pradeep Ravikumar

Machine Learning 10-701
Mar 1, 2017

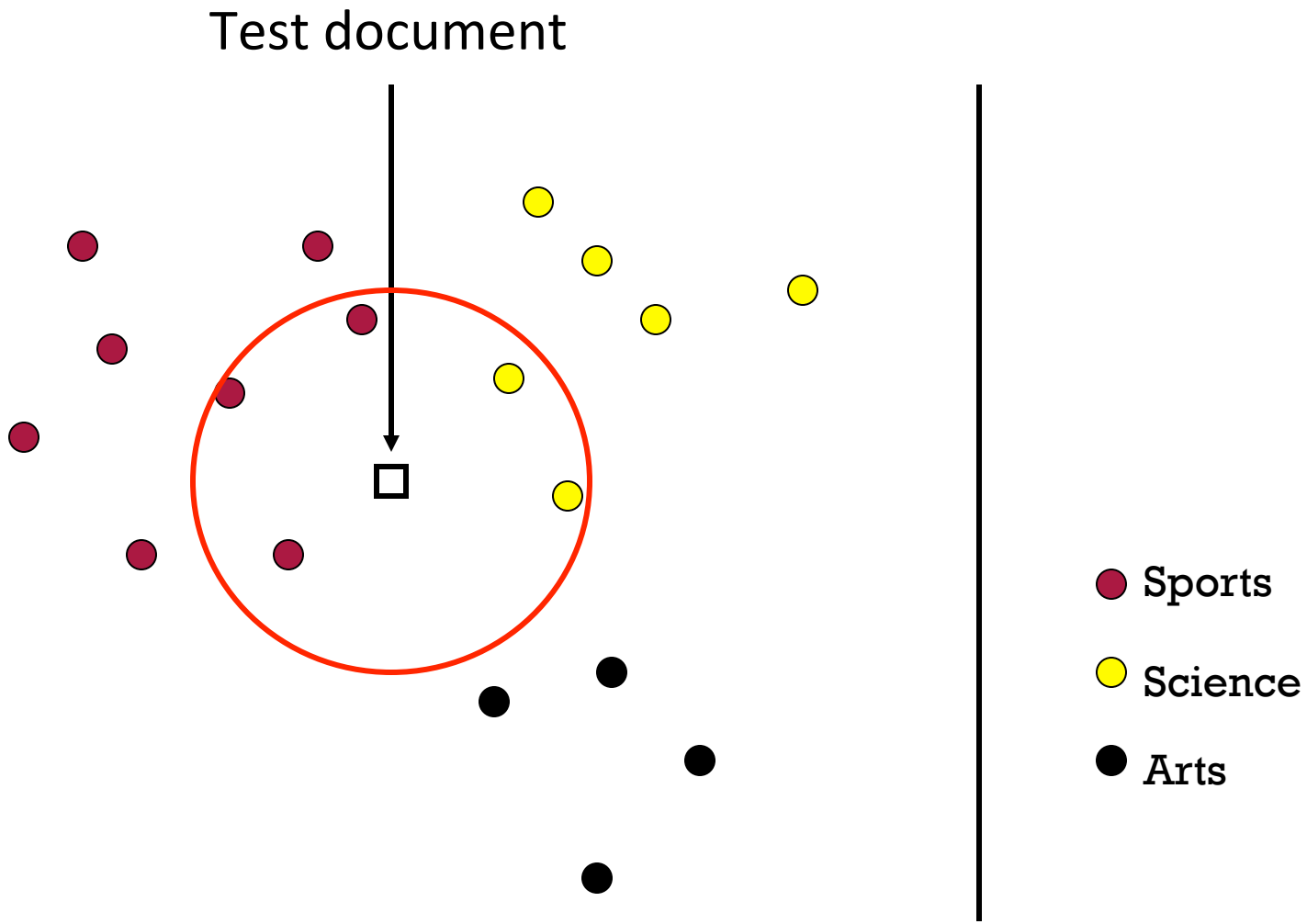
k-NN classifier



k-NN classifier

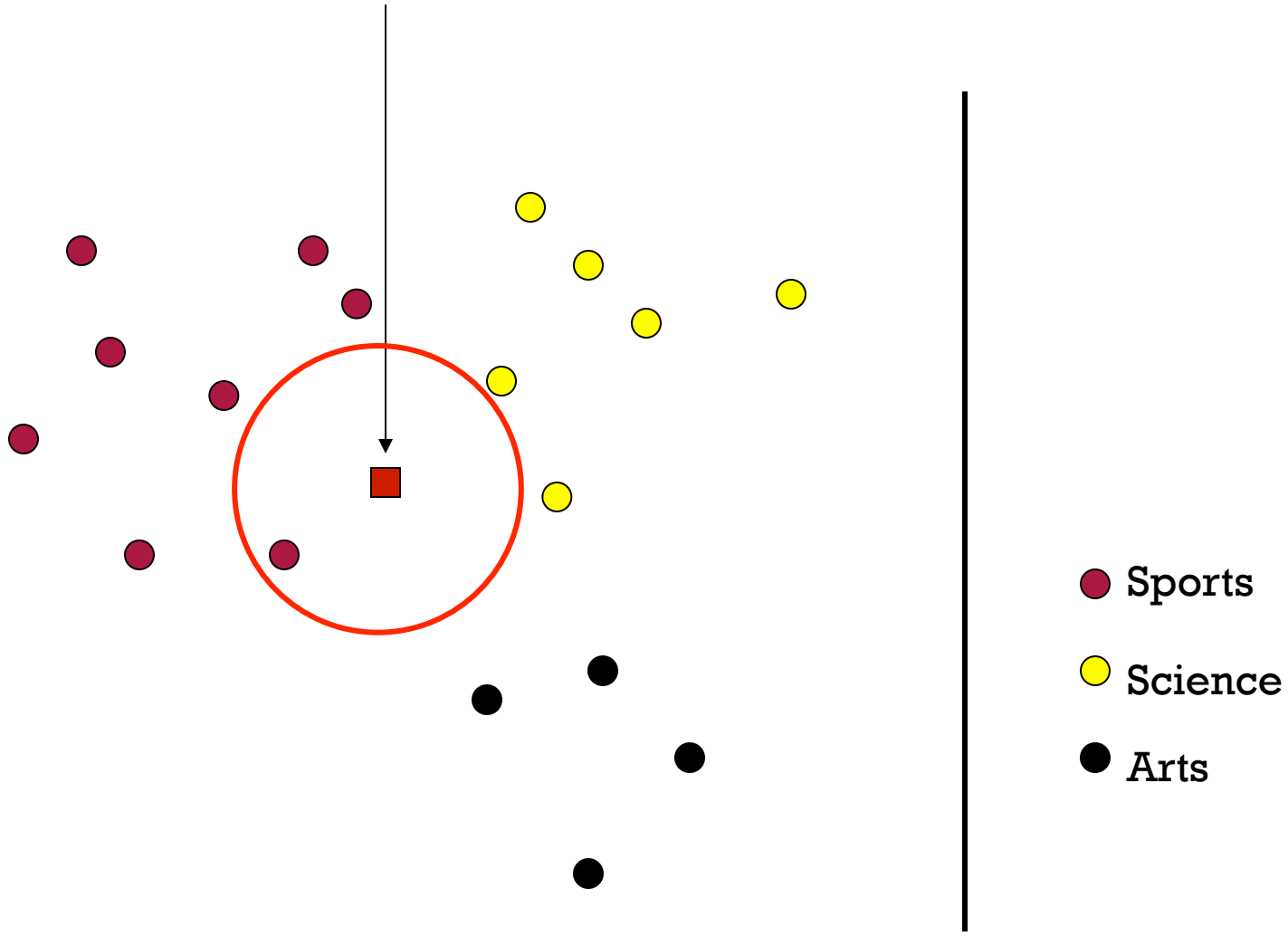


k-NN classifier (k=5)

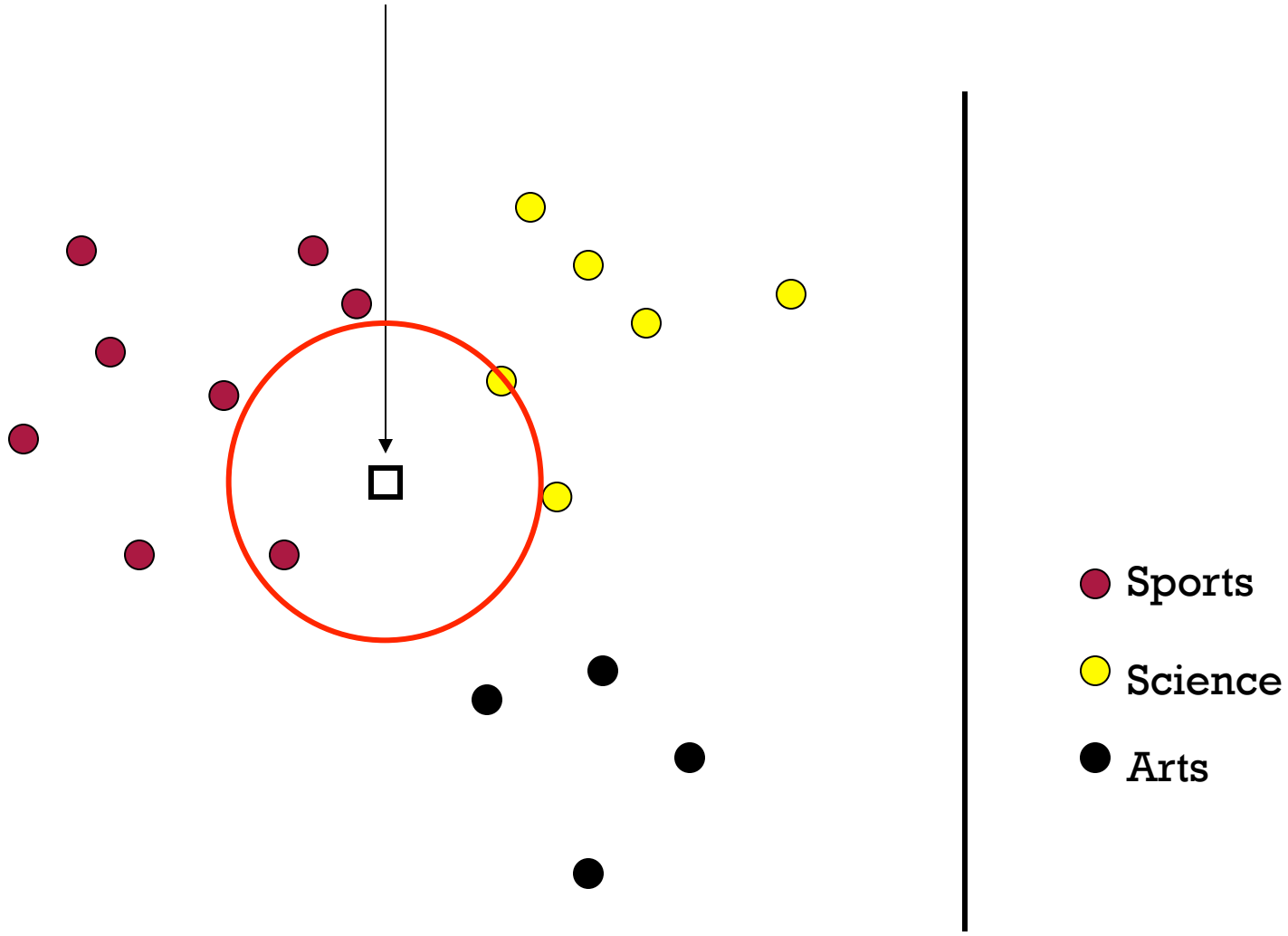


What should we predict? … Average? Majority? Why?

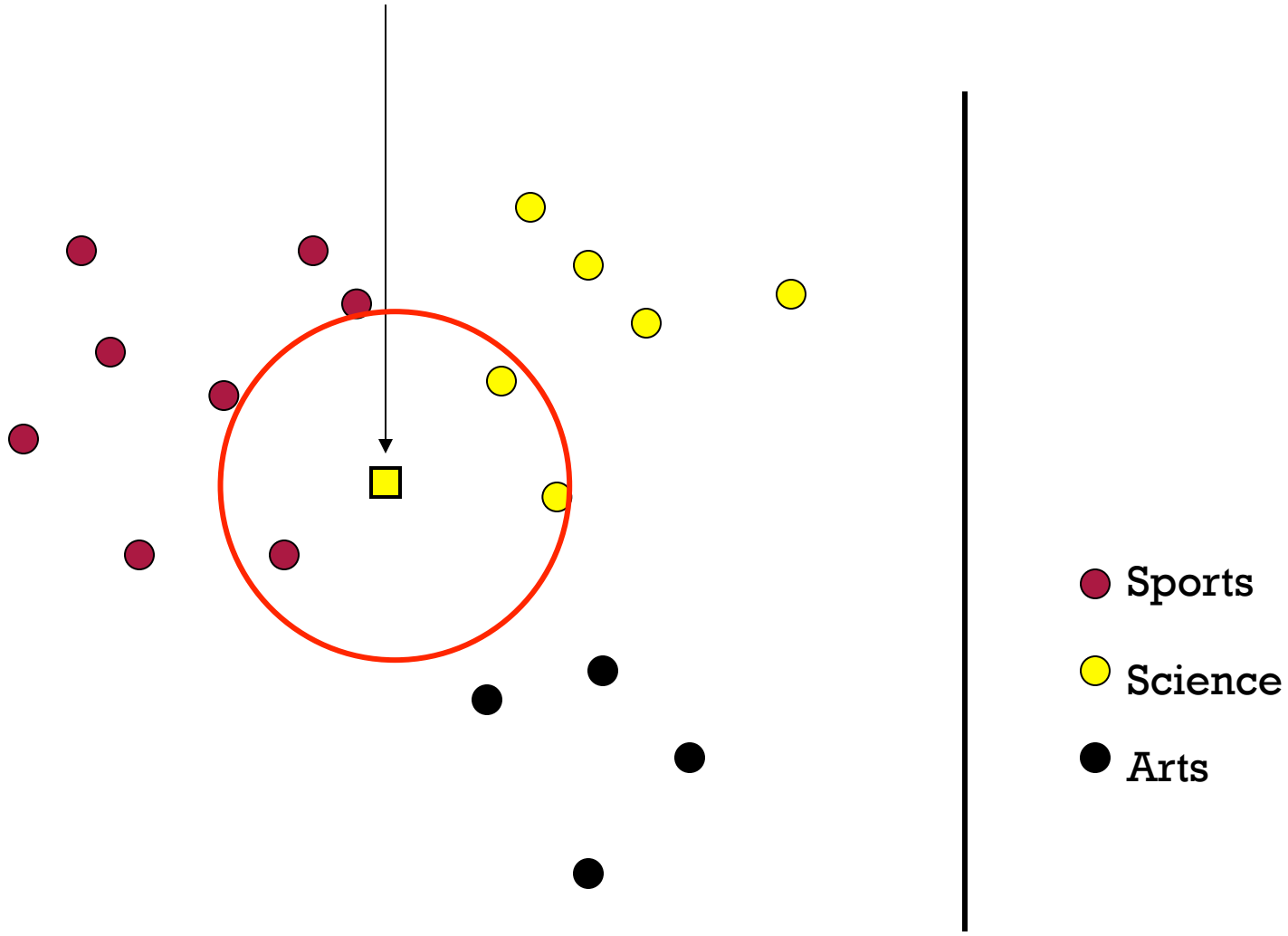
1-Nearest Neighbor (kNN) classifier



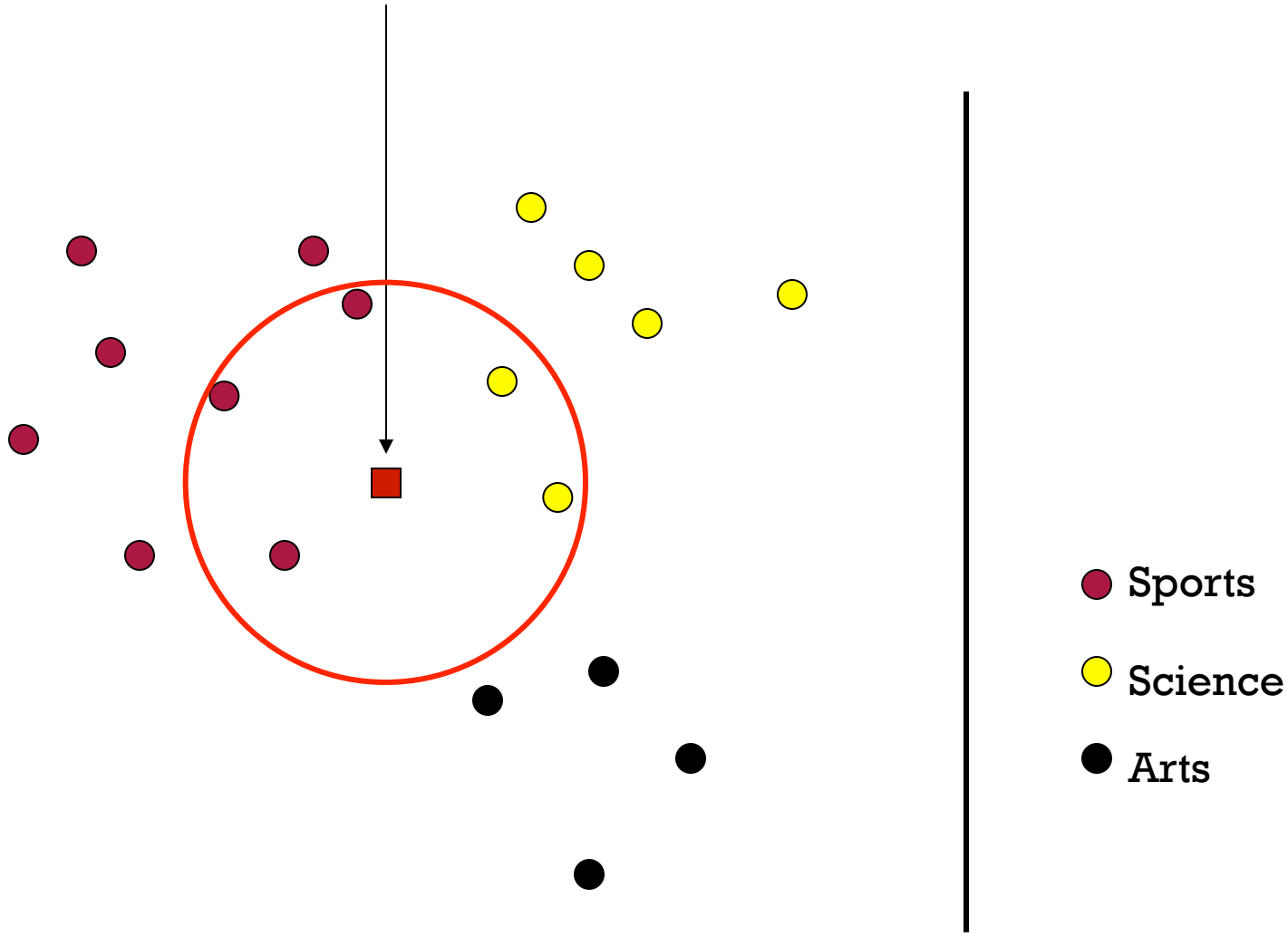
2-Nearest Neighbor (kNN) classifier



3-Nearest Neighbor (kNN) classifier

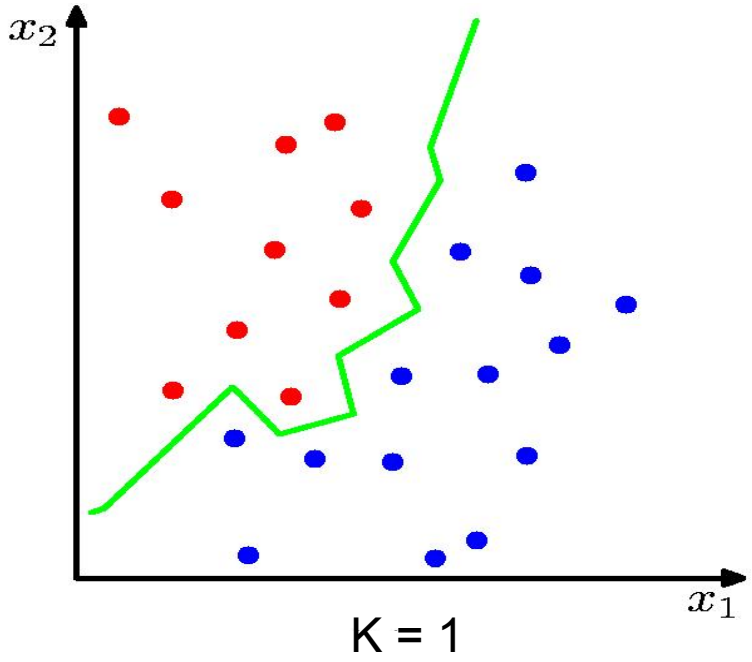


5-Nearest Neighbor (kNN) classifier



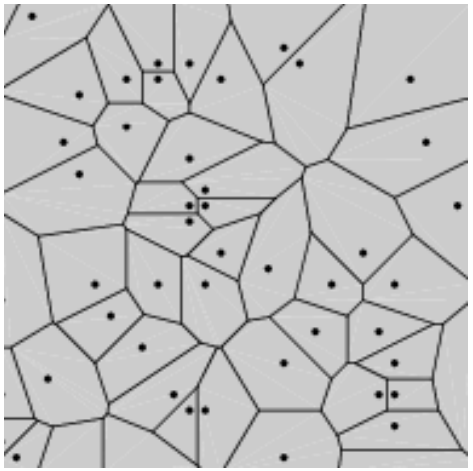
What is the best k?

1-NN classifier decision boundary



K = 1

Voronoi Diagram



As k increases, boundary becomes smoother (less jagged).

What is the best k?

Approximation vs. Stability Tradeoff

- Larger K => predicted label is more stable
- Smaller K => predicted label can approximate best classifier well

Non-parametric methods

Aka Instance-based/Memory-based/Lazy learners

- Decision Trees
- k-Nearest Neighbors

Parametric methods

- Assume some model (Gaussian, Bernoulli, Multinomial, logistic, network of logistic units, Linear, Quadratic) with fixed number of parameters
  - Gaussian Bayes, Naïve Bayes, Logistic Regression, Perceptron, Neural Networks

- Estimate parameters ($\mu, \sigma^2, \theta, w, \beta$) using MLE/MAP and plug in

- **Pro** – need few data points to learn parameters
- **Con** – Strong distributional assumptions, not satisfied in practice

Non-Parametric methods

- Typically don't make any distributional assumptions
- As we have more data, we should be able to learn more complex models
- Let number of parameters scale with number of training data

- Some nonparametric methods
  - Decision Trees
  - k-NN (k-Nearest Neighbor) Classifier

Summary

- Parametric vs Nonparametric approaches
  - Nonparametric models place very mild assumptions on the data distribution and provide good models for complex data

    Parametric models rely on very strong (simplistic) distributional assumptions

  - Nonparametric models requires storing and computing with the entire data set.

    Parametric models, once fitted, are much more efficient in terms of storage and computation.