

Statistics for Machine Learning and Big Data

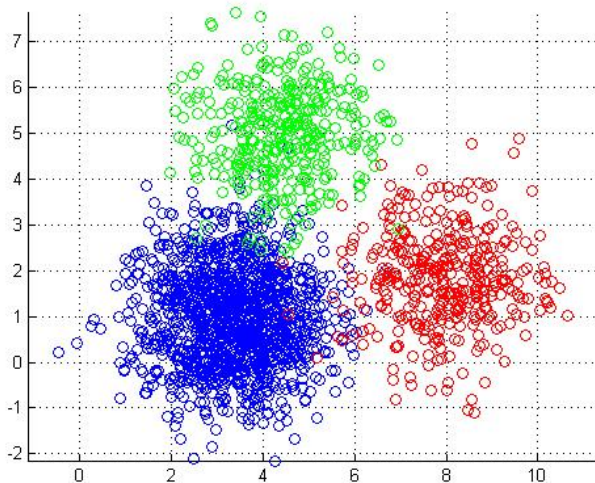
An Introduction

Part 5: Decision algorithms

Jeff Abrahamson

7 April 2015

k -means



Code time

Discussion

When should we use k -means vs logistic regression?

SVM

Support Vector Machines

- Goal: optimal separating hyperplane
- aka: Large Margin Classifier

Strategies

At the beginning, one tends to do this:

- Transform data for SVM solver
- Randomly try a few kernels and parameters
- Test

Strategies

A better strategy:

- Transform
- Scale data
- Consider linear, Gaussian, or RBF kernels
- Use cross validation to find the best C and γ
- Test

Strategies

Categories usually work better as binary fields than as enums.

Break

Questions?

`purple.com/talk-feedback`