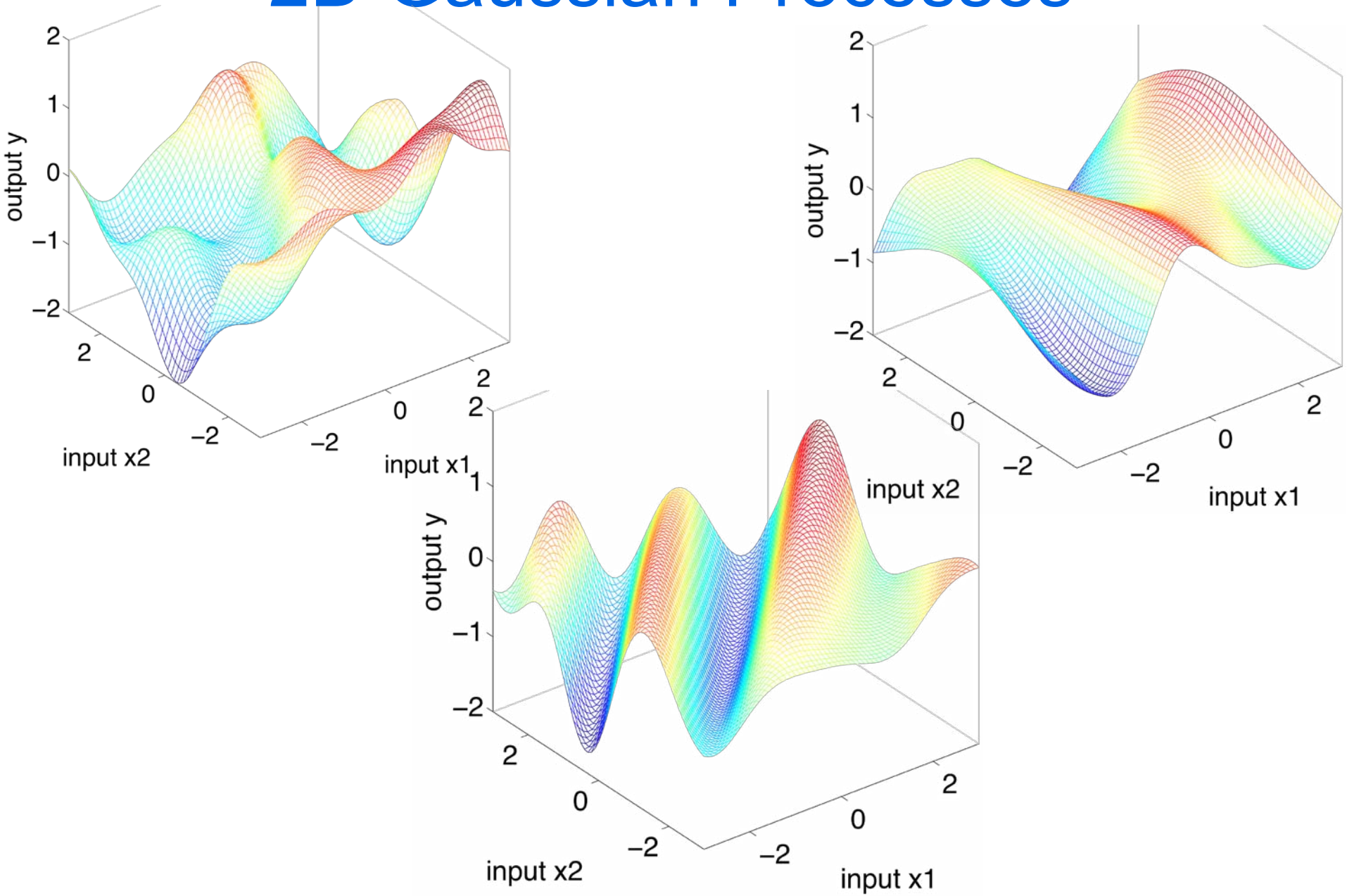
Introduction to Machine Learning

Brown University CSCI 1950-F, Spring 2011
Prof. Erik Sudderth

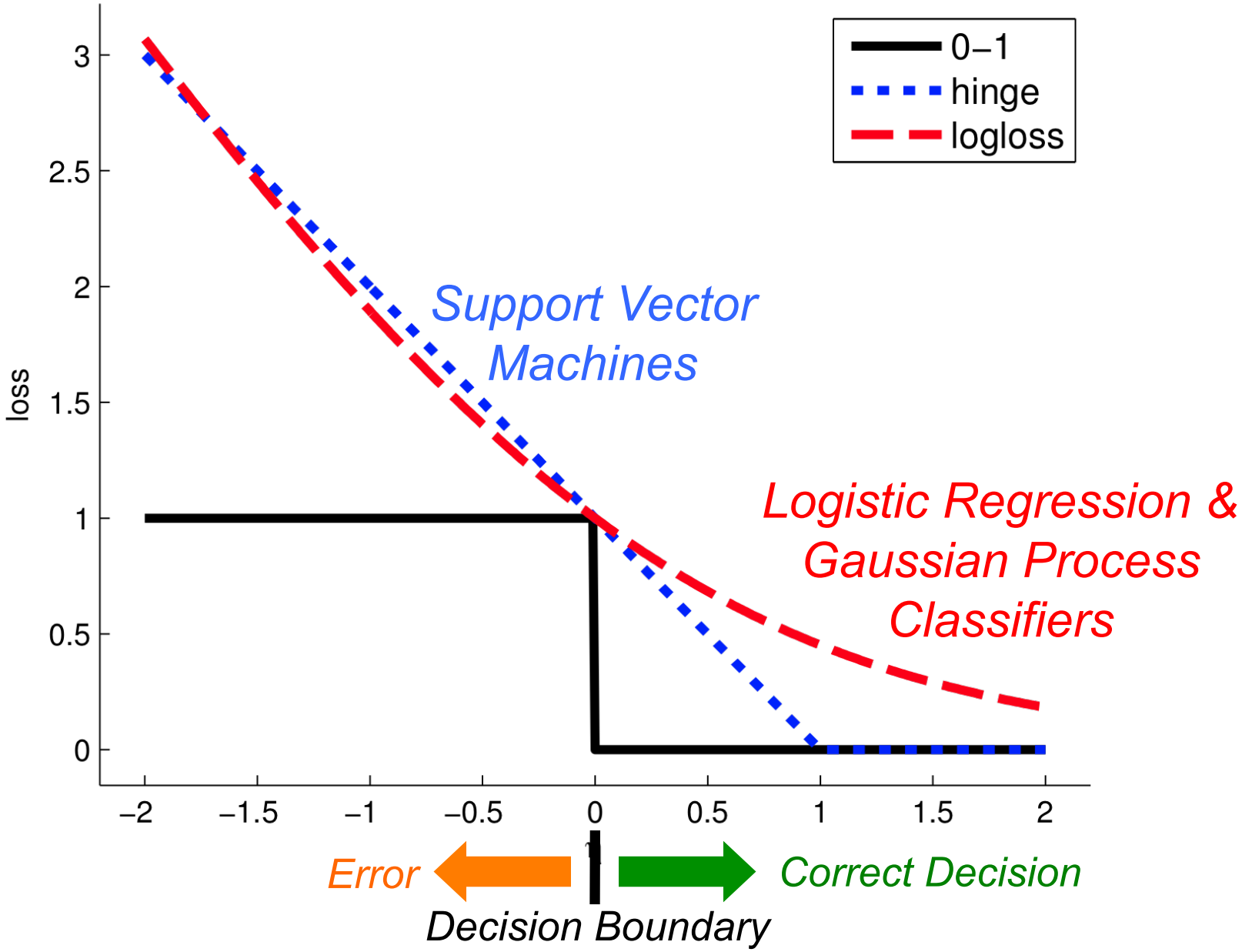
Lecture 17: Support Vector Machines, Clustering, K-Means

Many figures courtesy Kevin Murphy's textbook,
Machine Learning: A Probabilistic Perspective

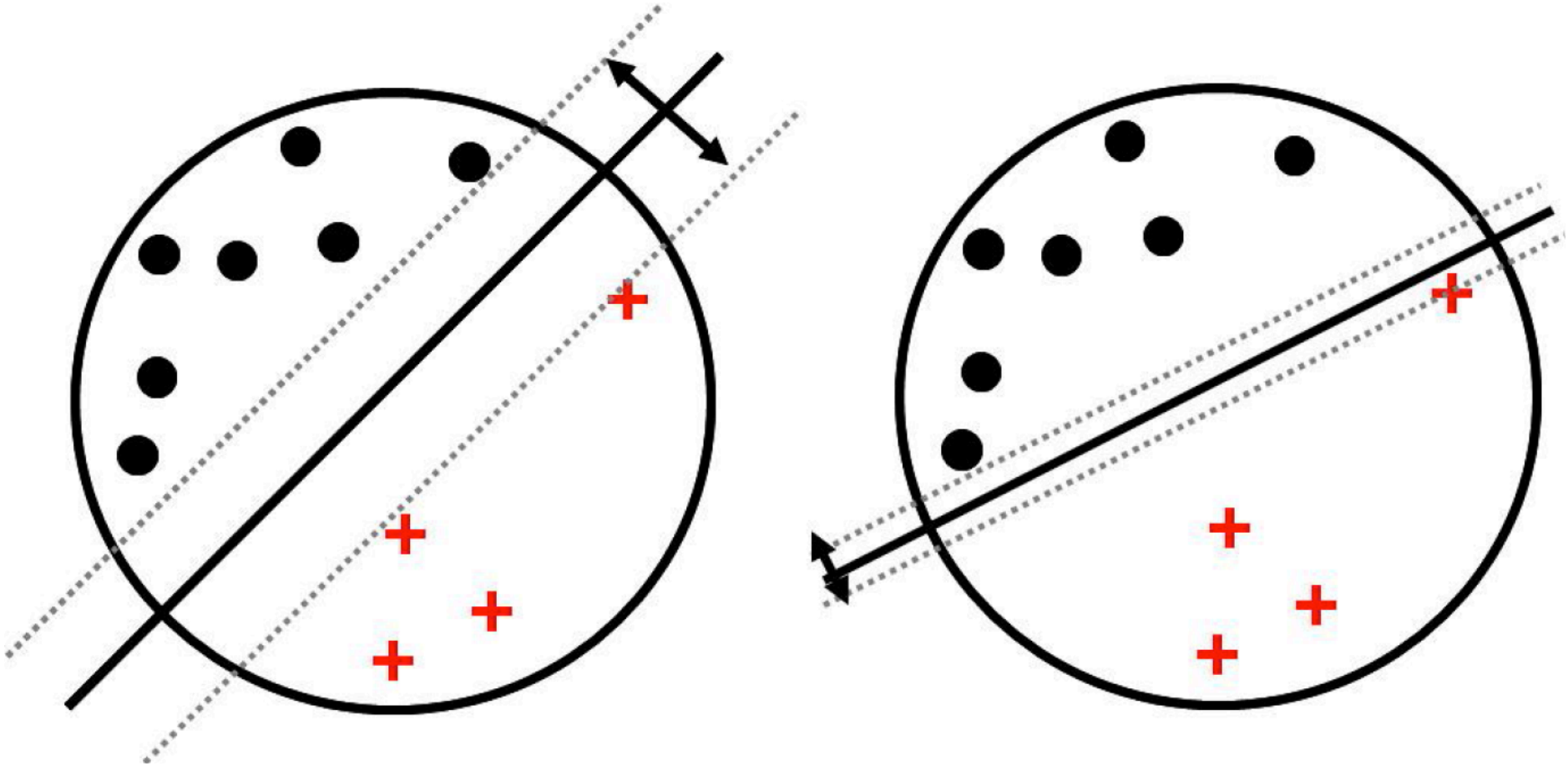
2D Gaussian Processes



Losses for Binary Classification

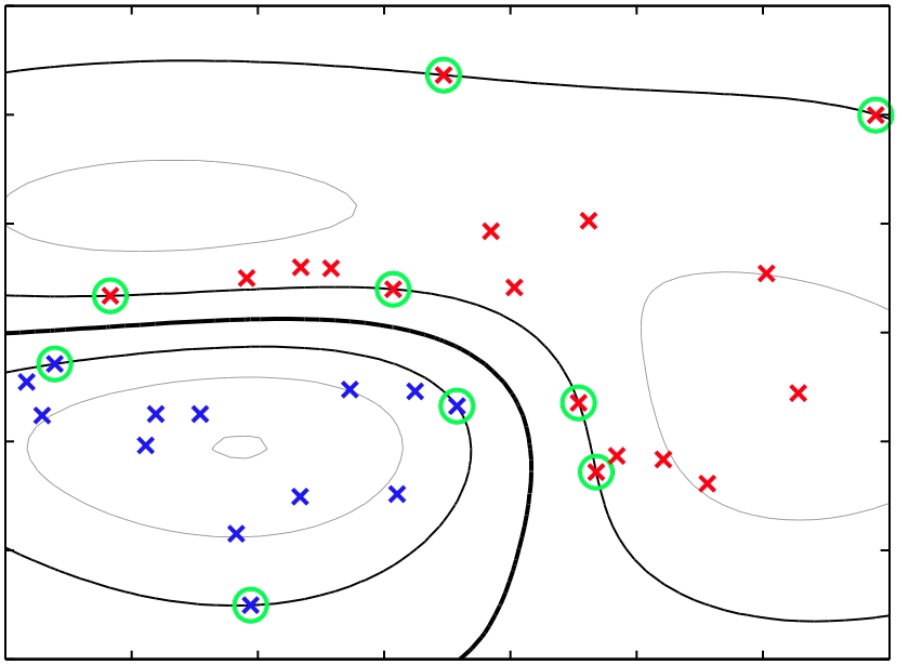


Maximum Margin Hyperplanes

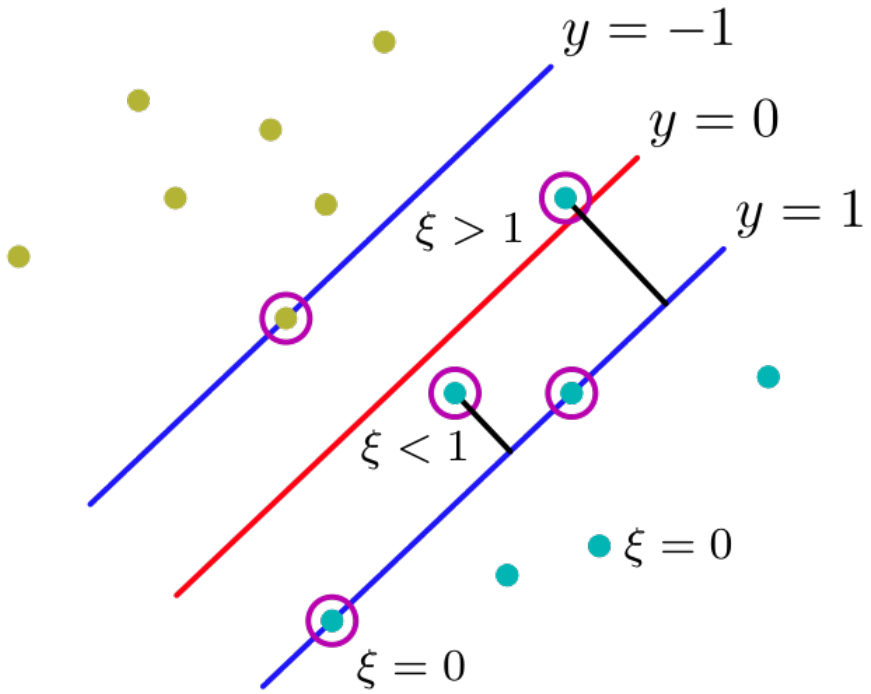


If multiple linear classifiers perfectly separate training data, which should I choose?

Support Vectors & Slack Variables



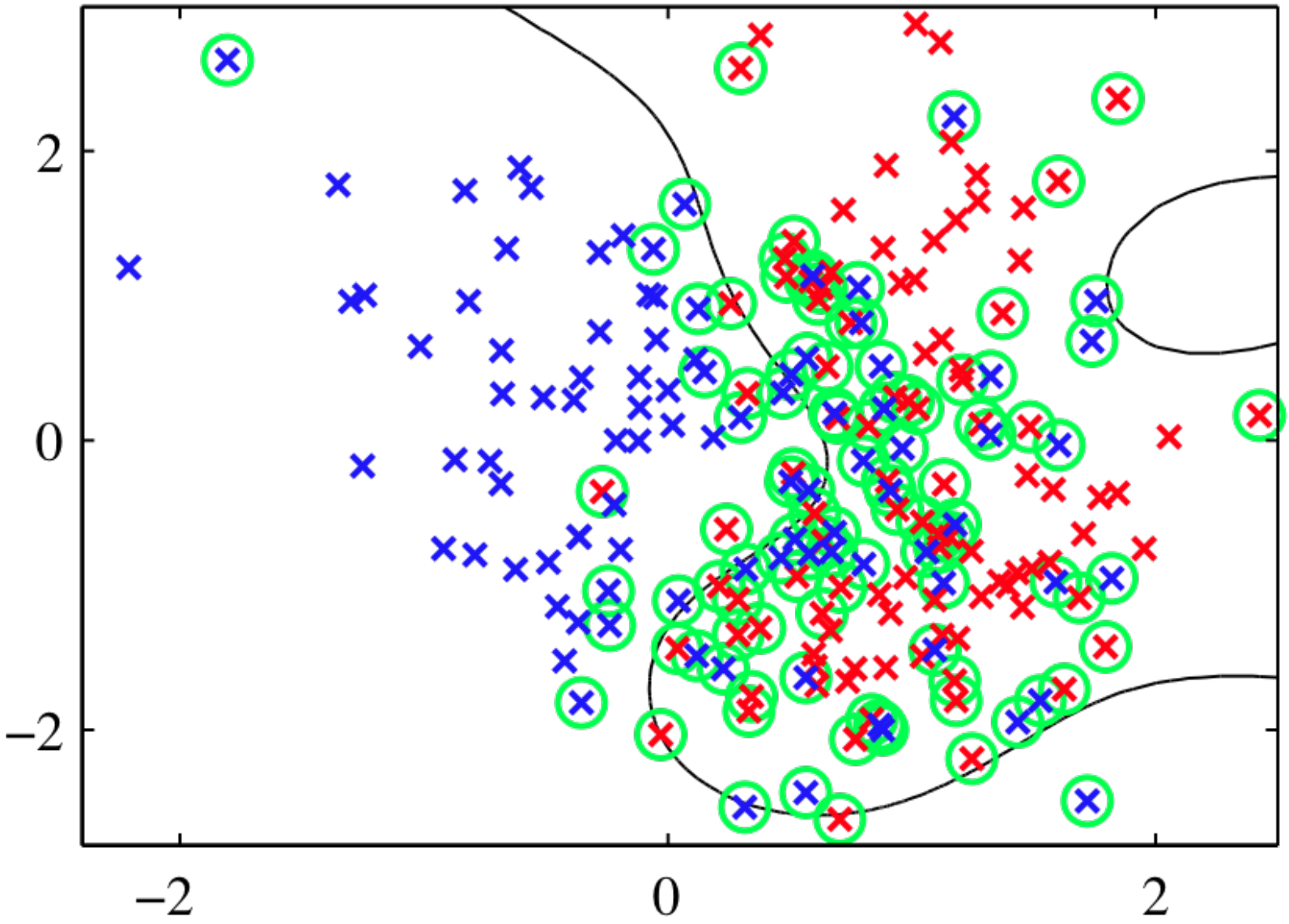
Support vectors (green) for data separable by radial basis function kernels, and non-linear margin boundaries



Linear decision boundary in feature space, where data violating margin have nonzero "slack variables"

C. Bishop, Pattern Recognition & Machine Learning

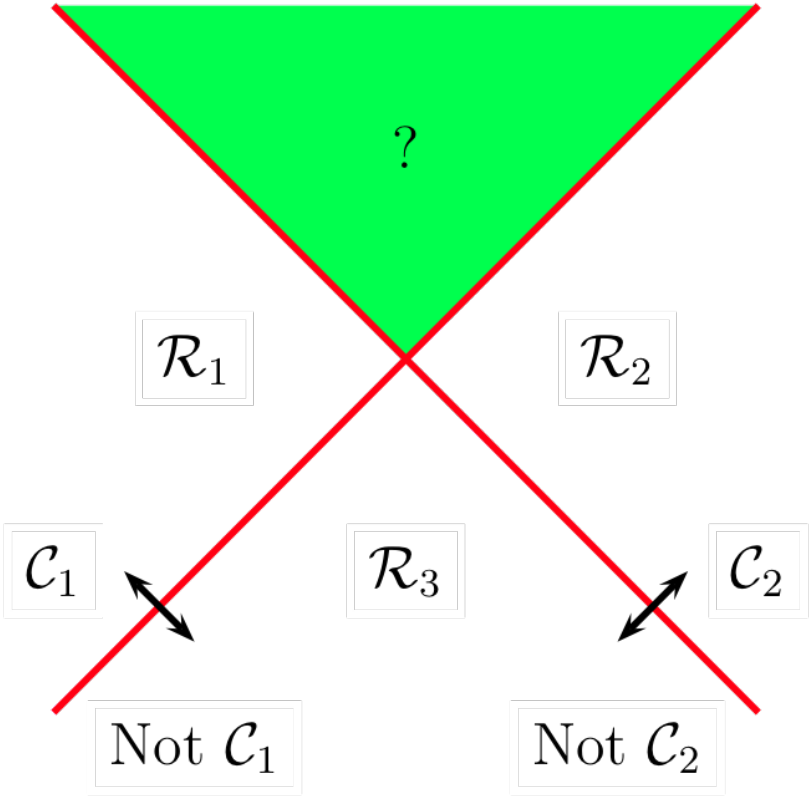
How Many Support Vectors?



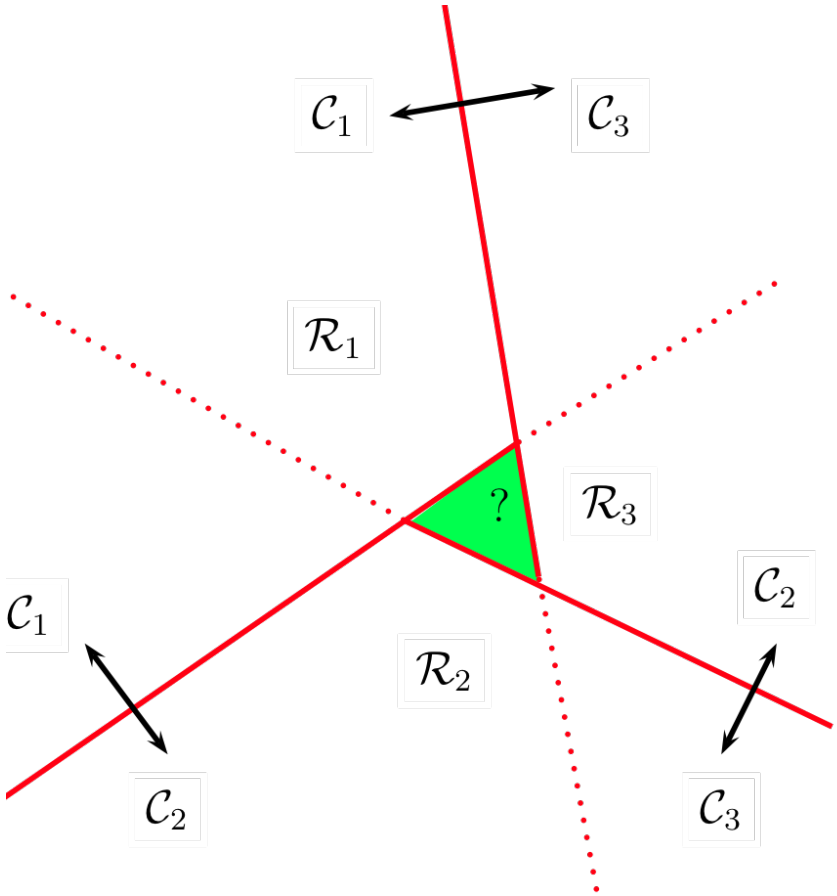
C. Bishop, Pattern Recognition & Machine Learning

Multiclass Support Vector Machines

*Complicated by the fact that binary SVM classifiers are **not** calibrated probabilistic models*



One versus Rest
(One versus All)



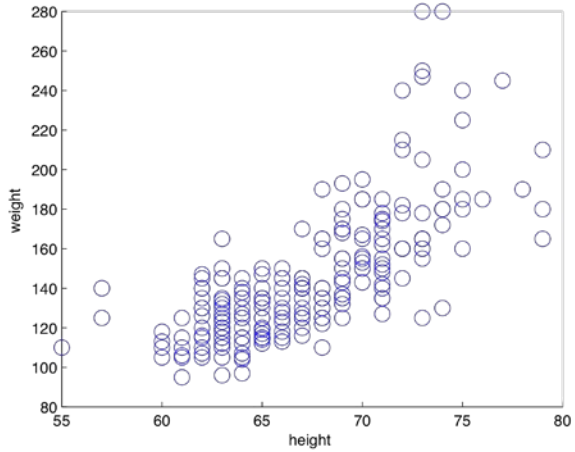
One versus One
(separate each pair of classes)

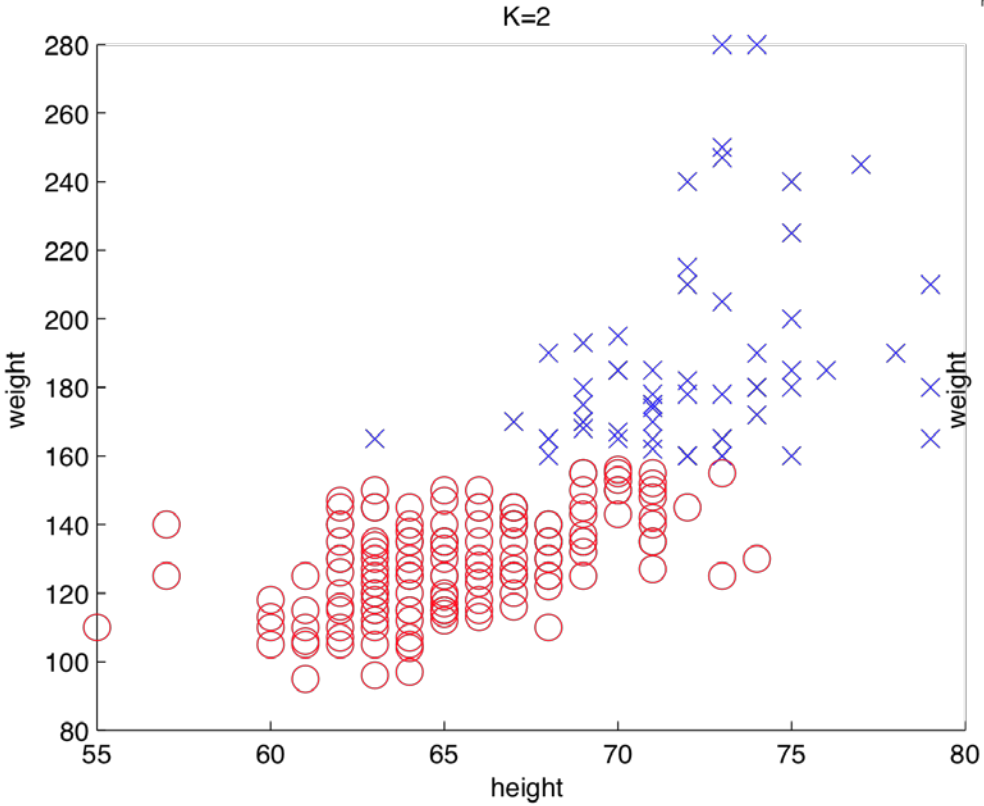
On to Unsupervised Learning

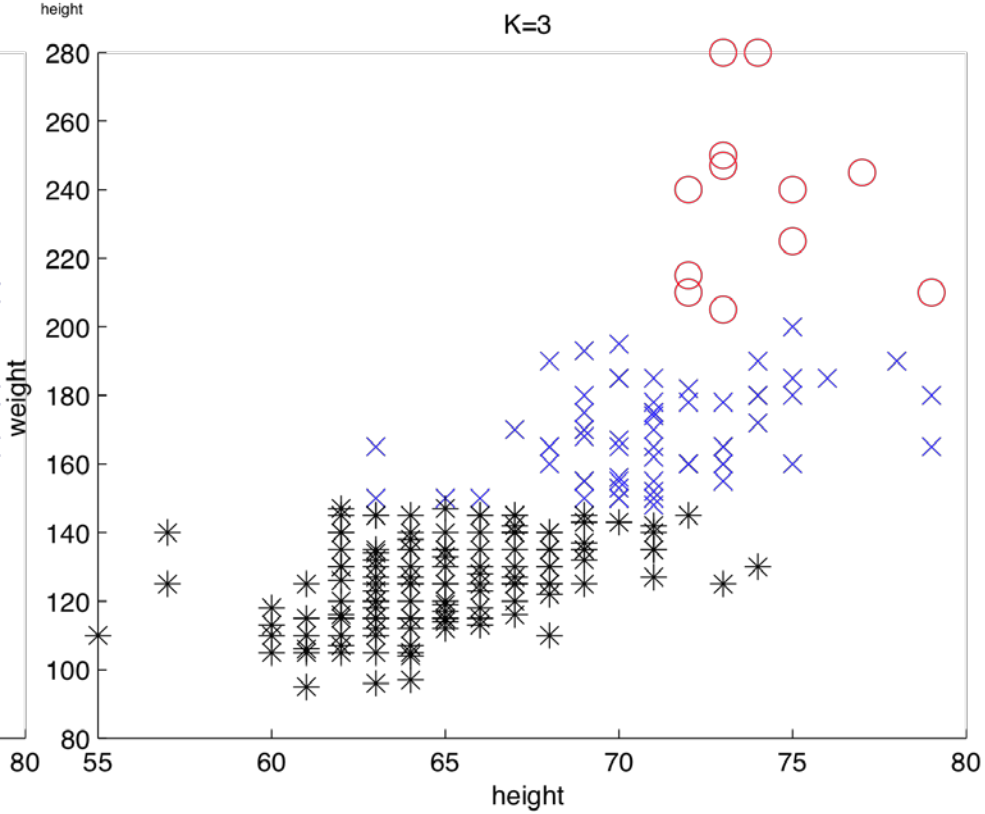
	Supervised Learning	*Unsupervised Learning*
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

- **Goal:** Infer label/response y given only features x
- **Classical:** Find latent variables y good for *compression* of x
- **Probabilistic learning:** Estimate parameters of joint distribution p(x,y) which *maximize marginal probability* p(x)

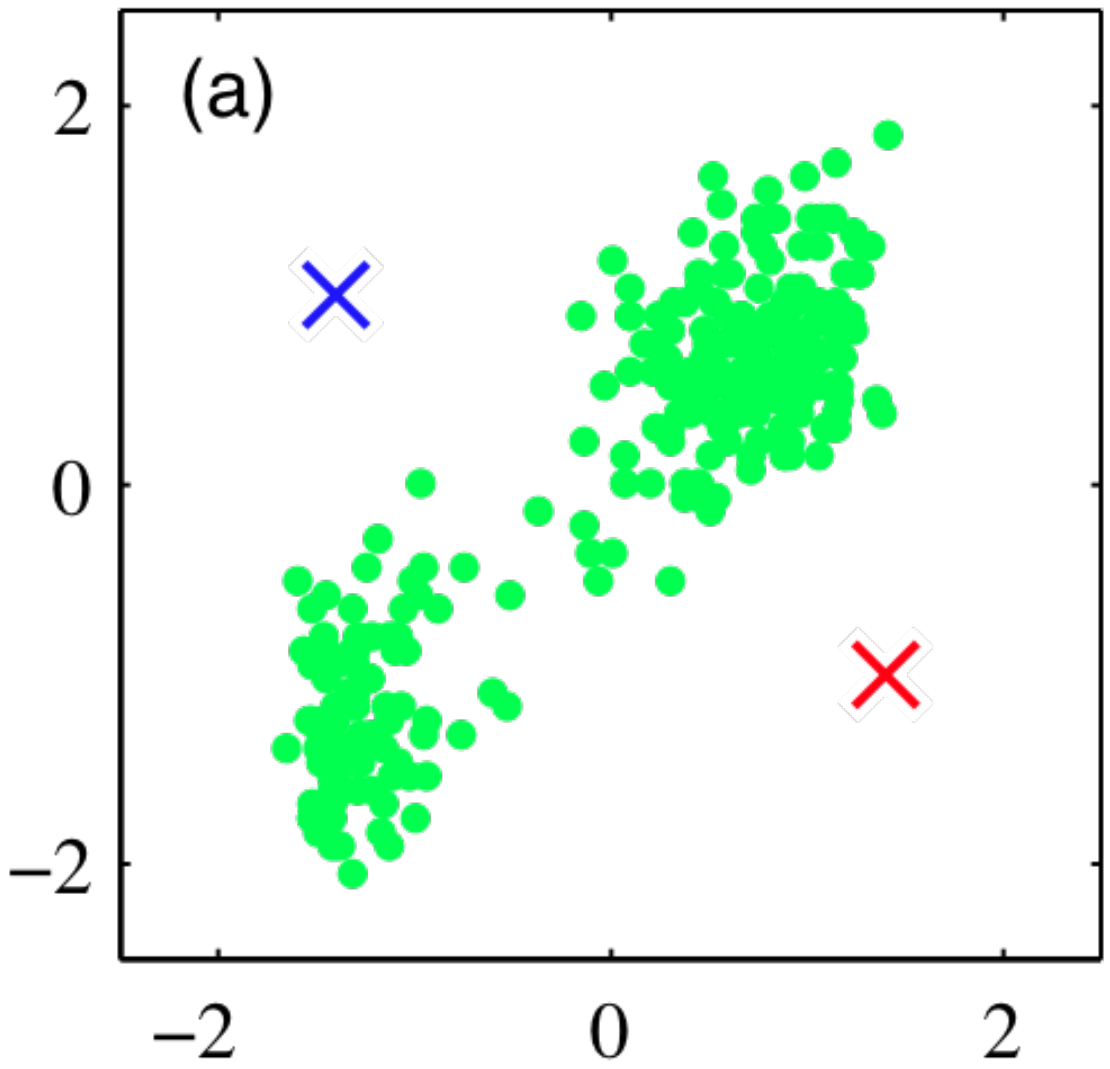
Clustering can be Ambiguous





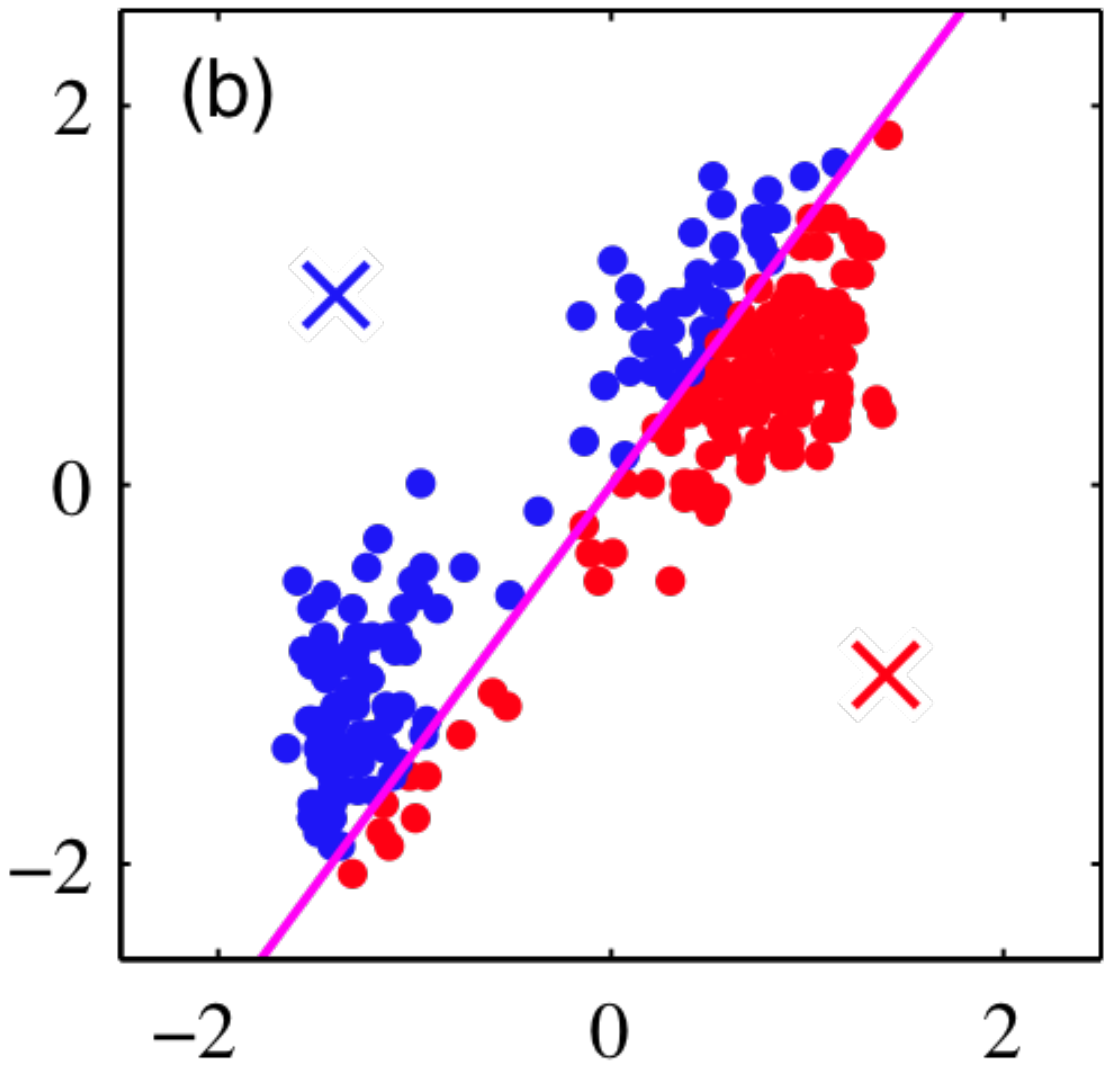


K-Means Algorithm



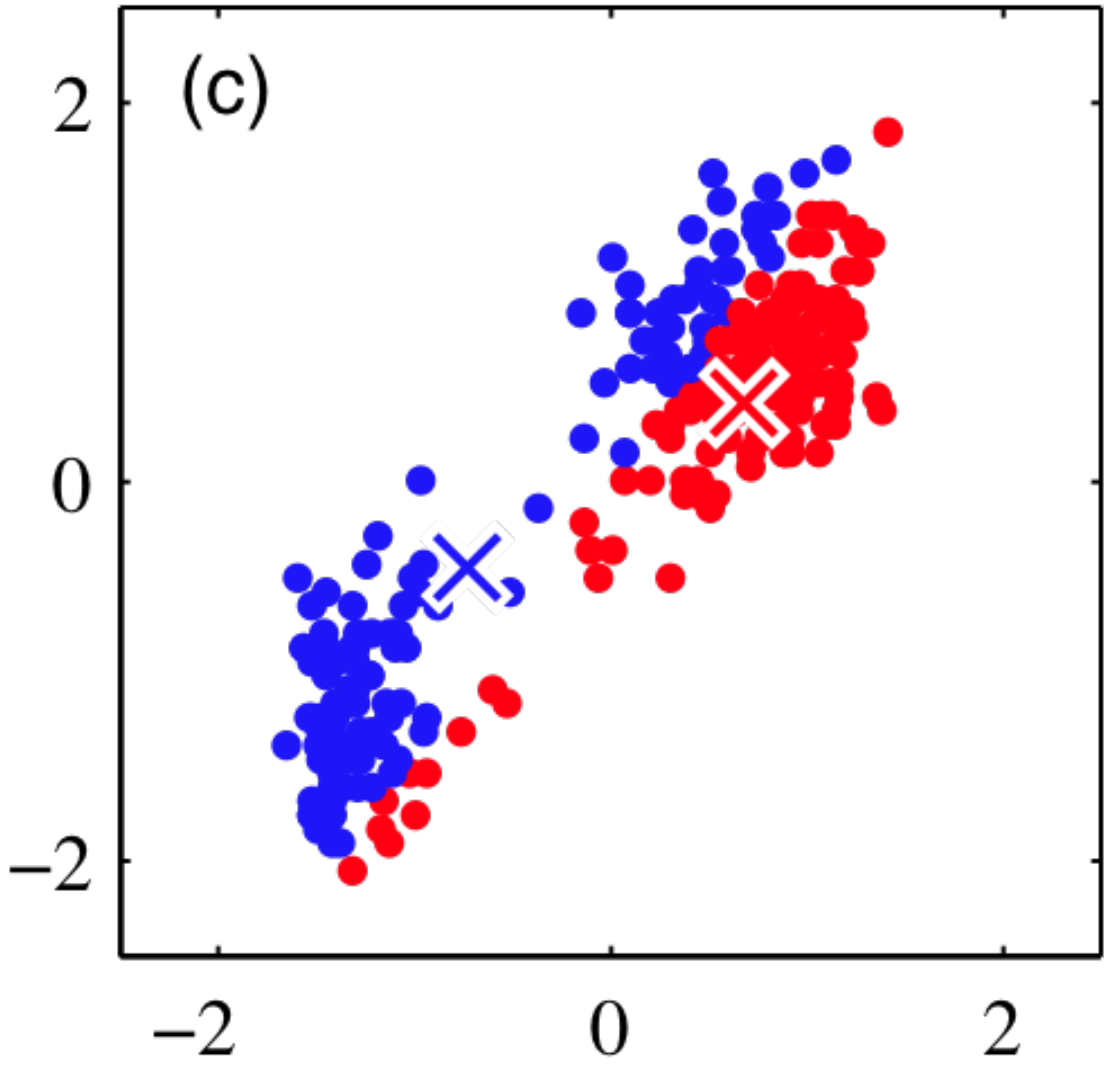
C. Bishop, Pattern Recognition & Machine Learning

K-Means Algorithm



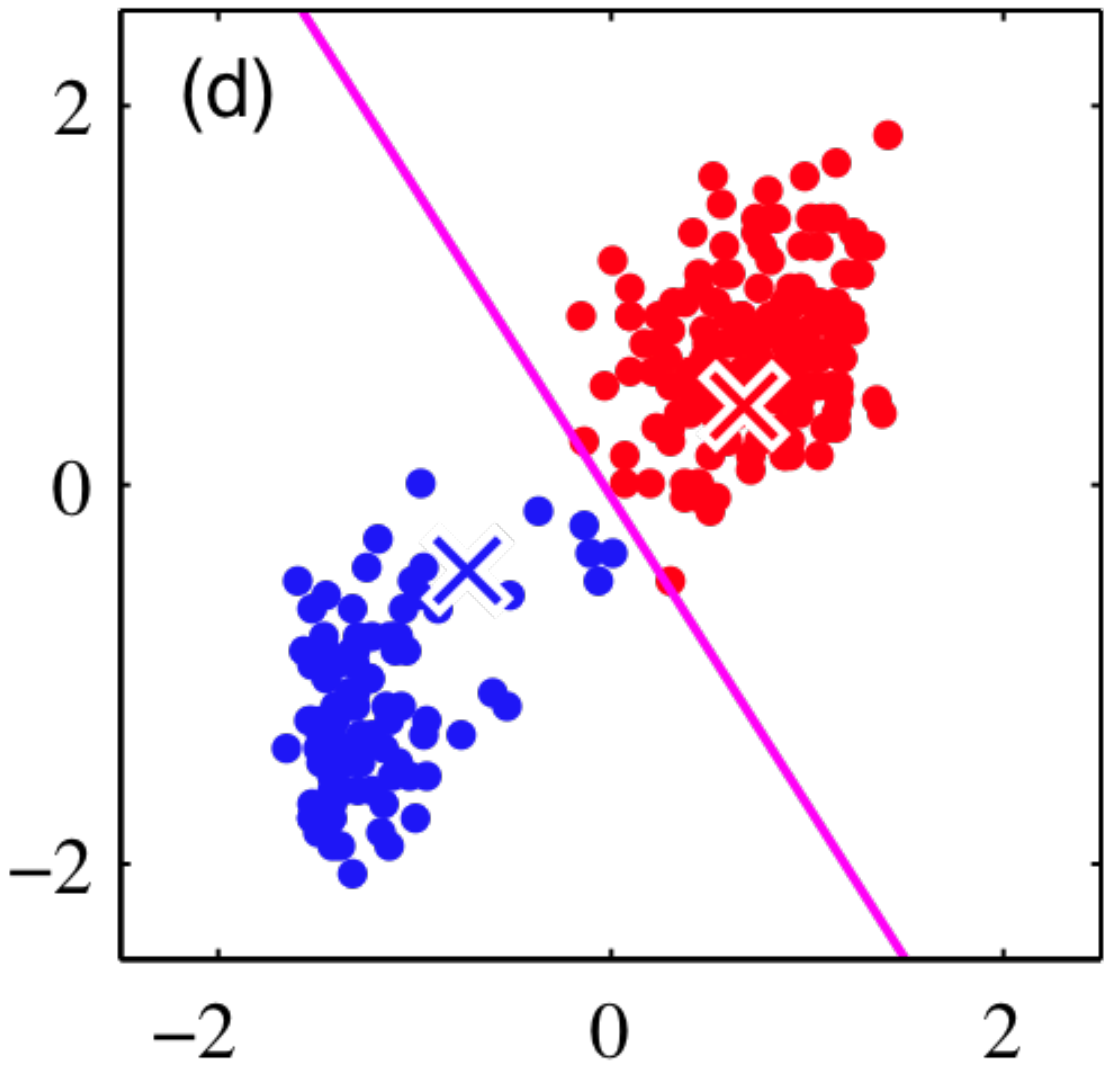
C. Bishop, Pattern Recognition & Machine Learning

K-Means Algorithm



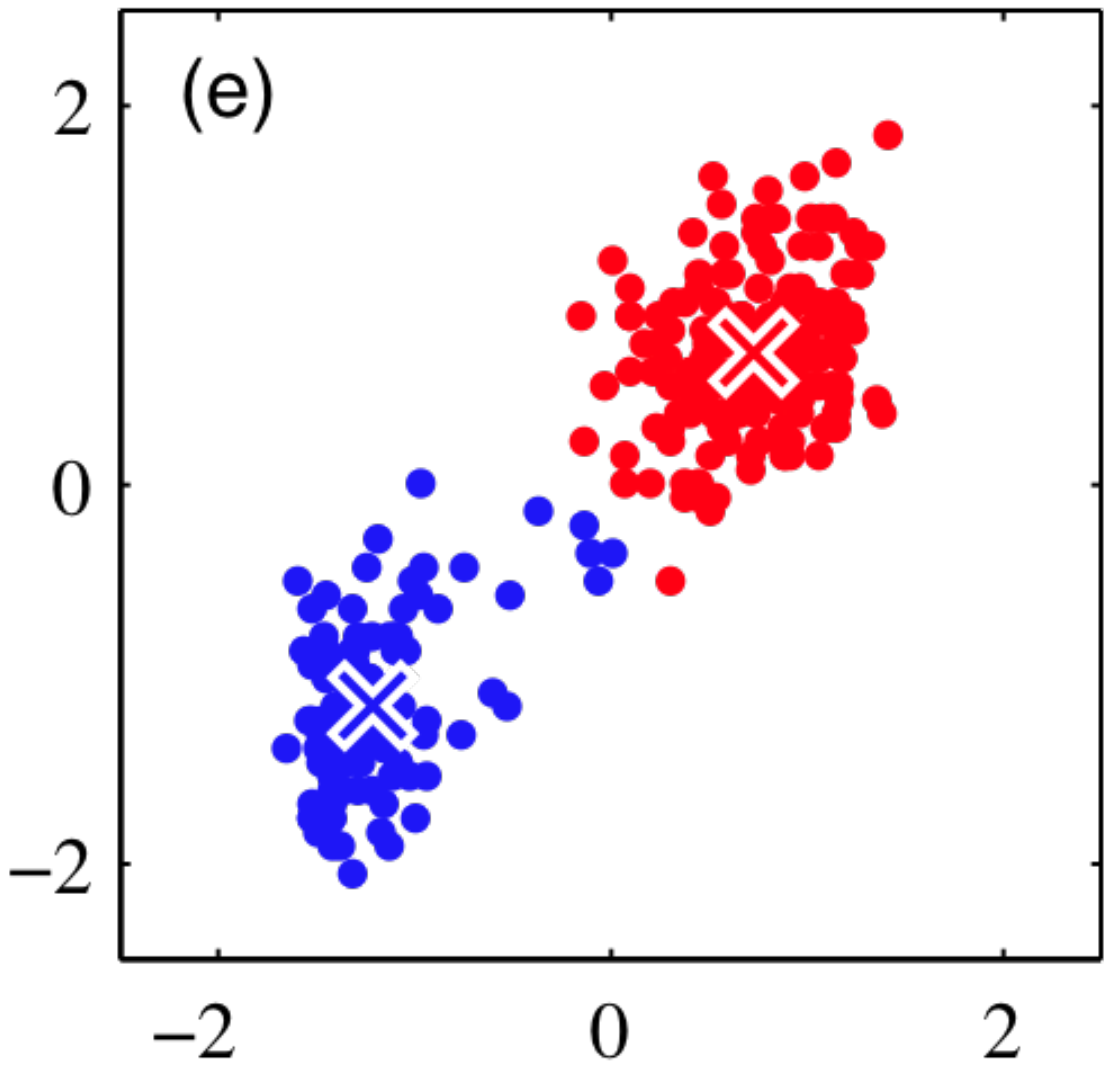
C. Bishop, Pattern Recognition & Machine Learning

K-Means Algorithm



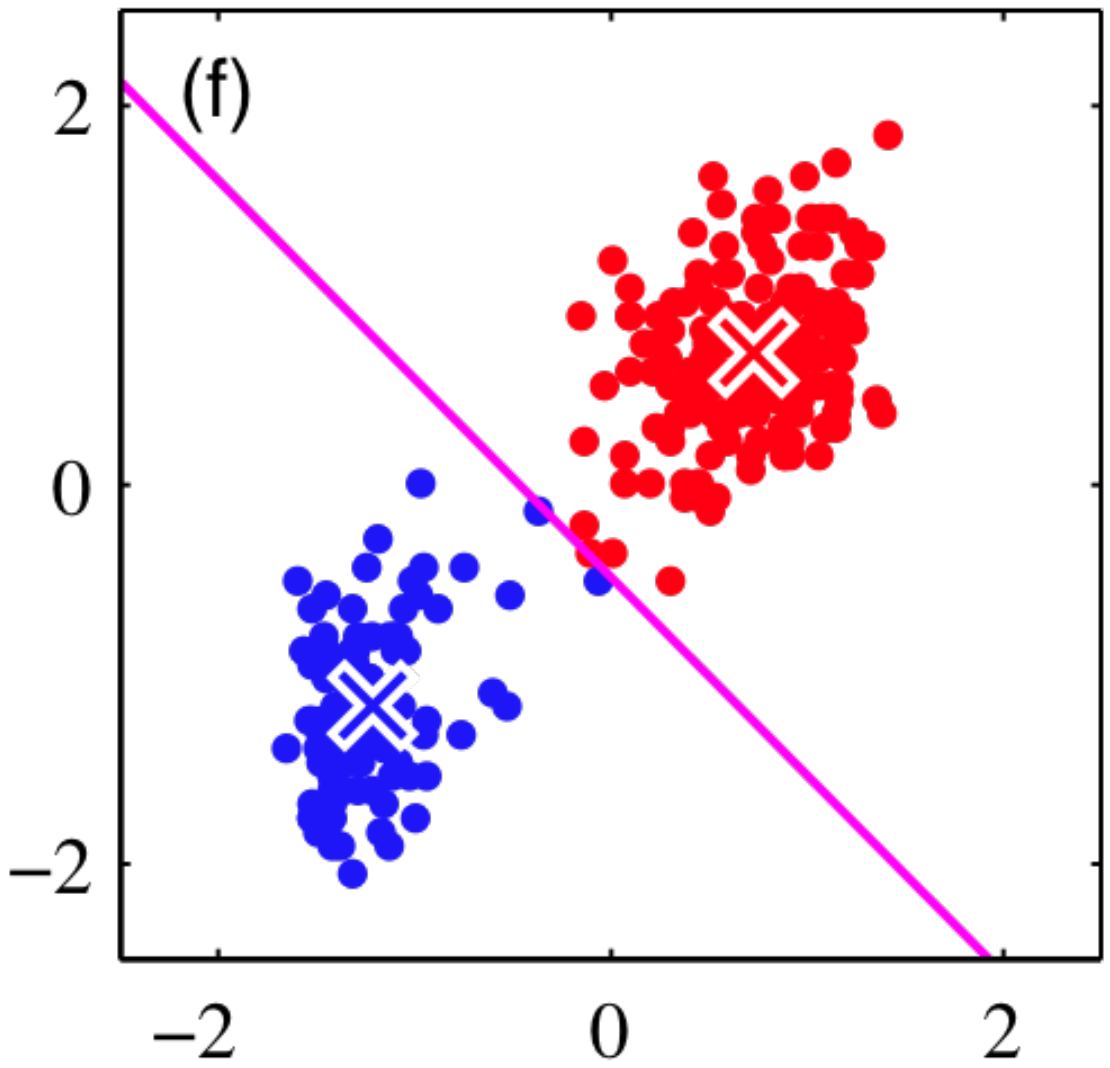
C. Bishop, Pattern Recognition & Machine Learning

K-Means Algorithm



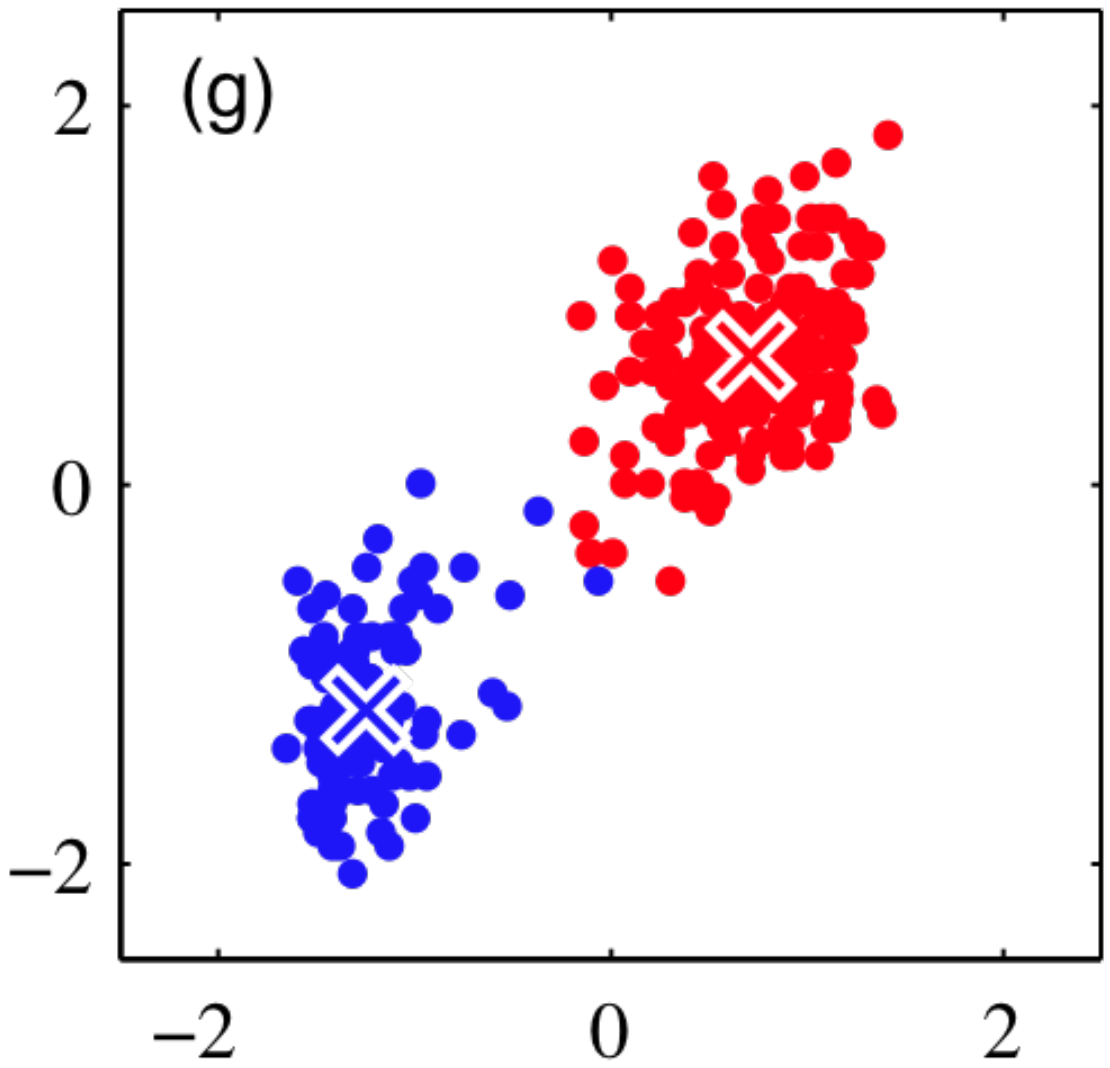
C. Bishop, Pattern Recognition & Machine Learning

K-Means Algorithm



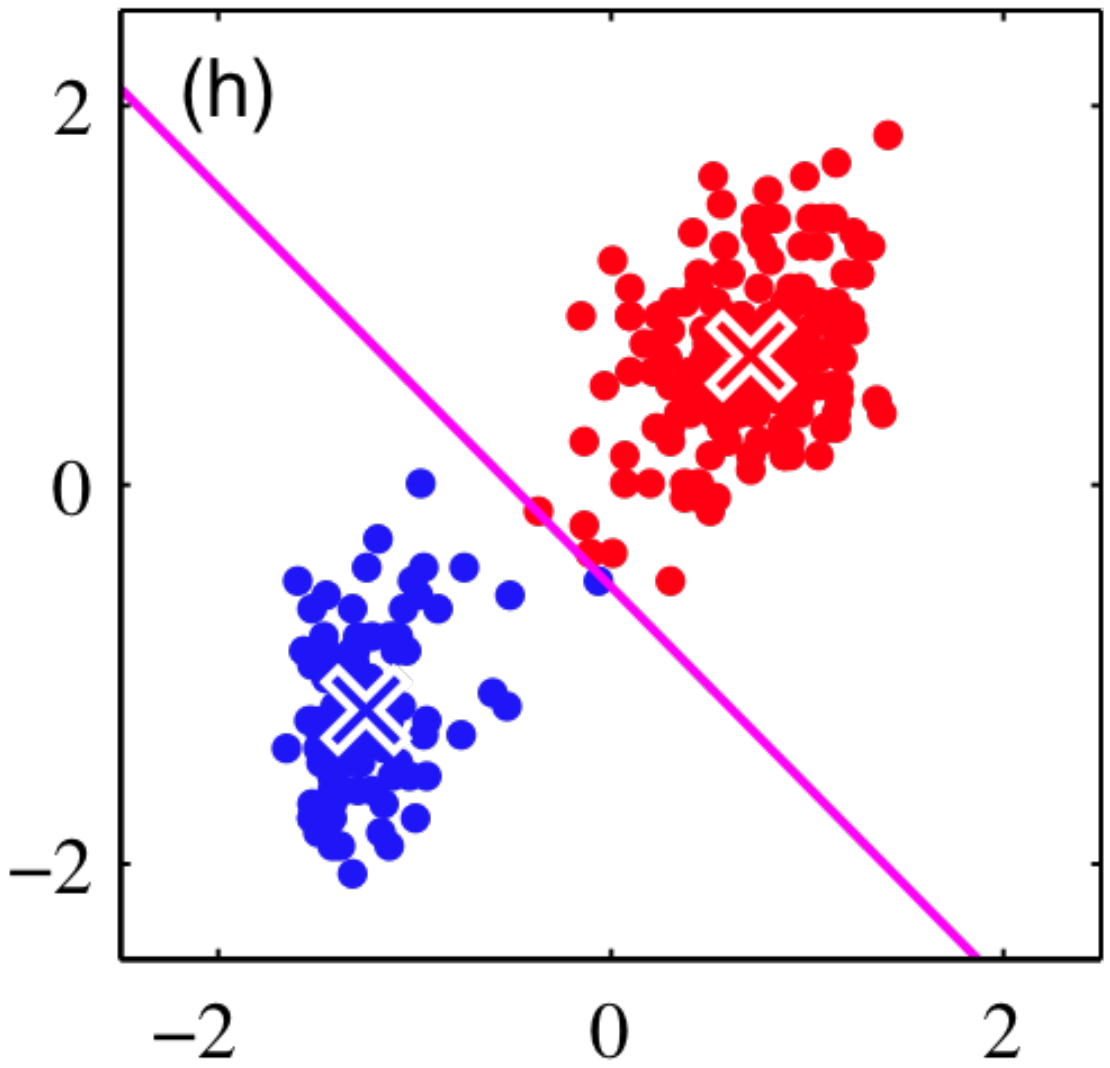
C. Bishop, Pattern Recognition & Machine Learning

K-Means Algorithm



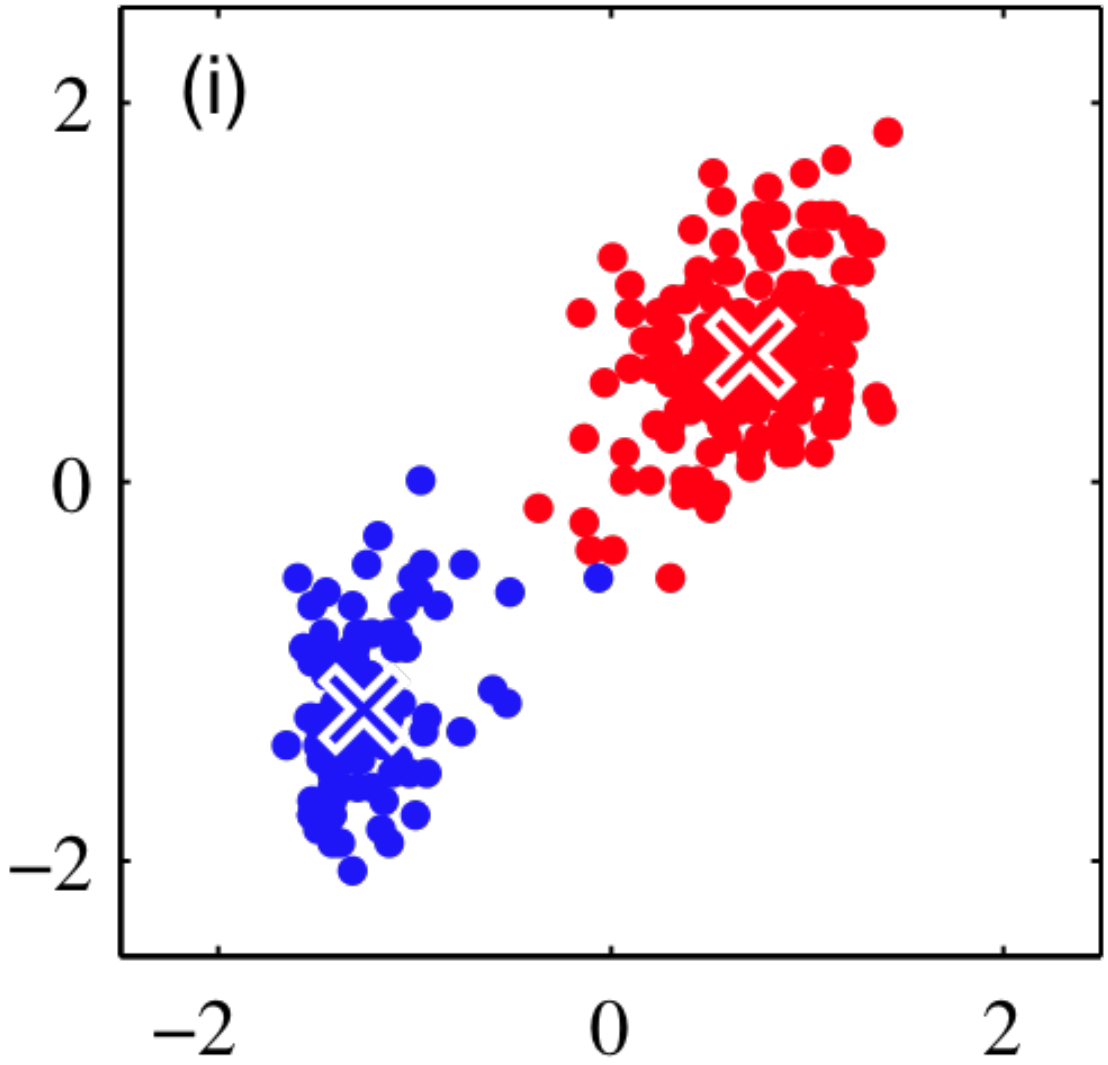
C. Bishop, Pattern Recognition & Machine Learning

K-Means Algorithm



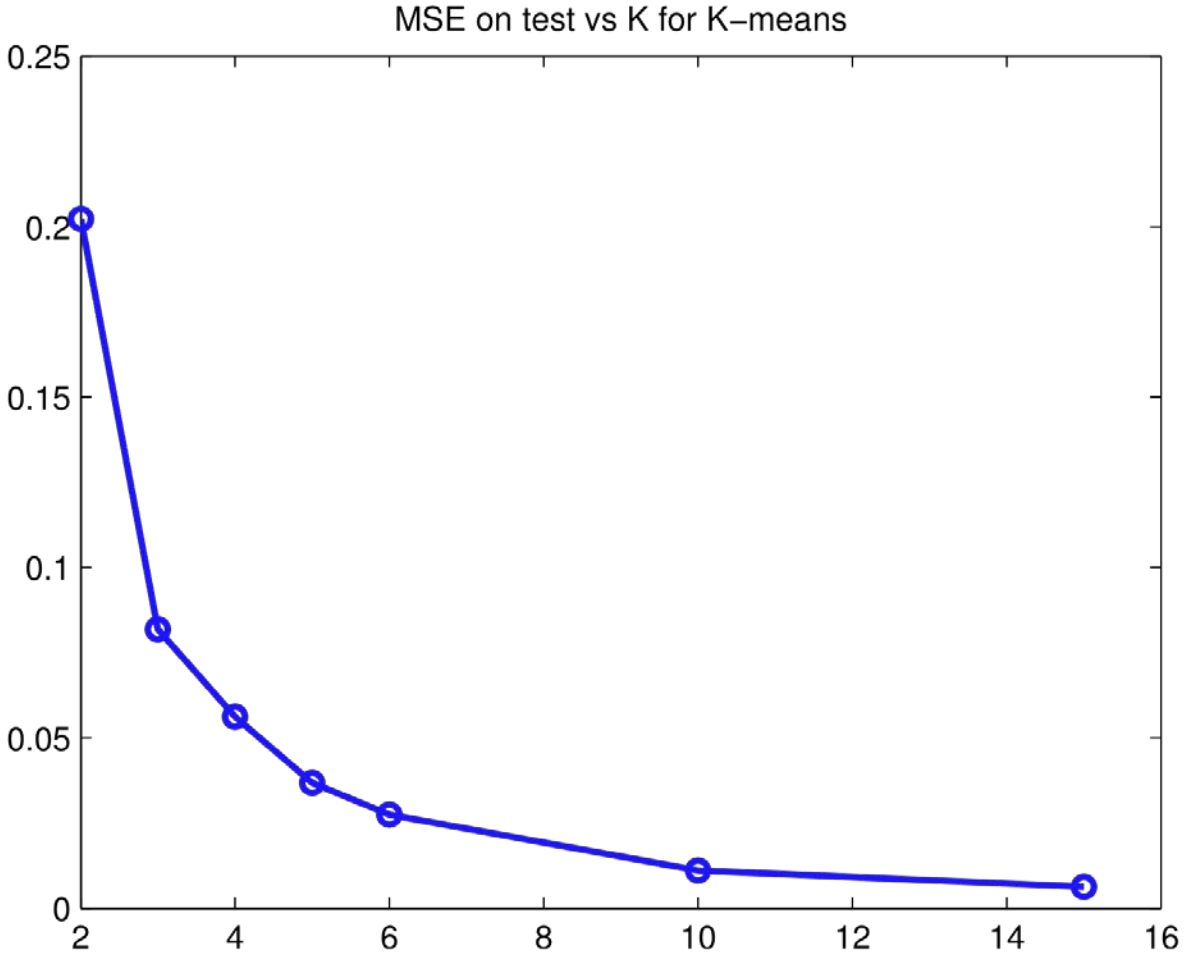
C. Bishop, Pattern Recognition & Machine Learning

K-Means Algorithm



C. Bishop, Pattern Recognition & Machine Learning

Test Error versus K



Data from a mixture of 3 1D Gaussian distributions

For compressing new data, more codewords is always better.

Cross-validation fails for unsupervised learning!