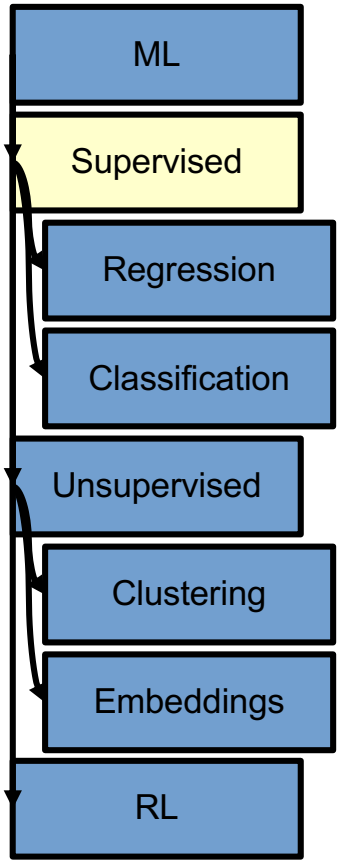
CS181: Introduction to Machine Learning

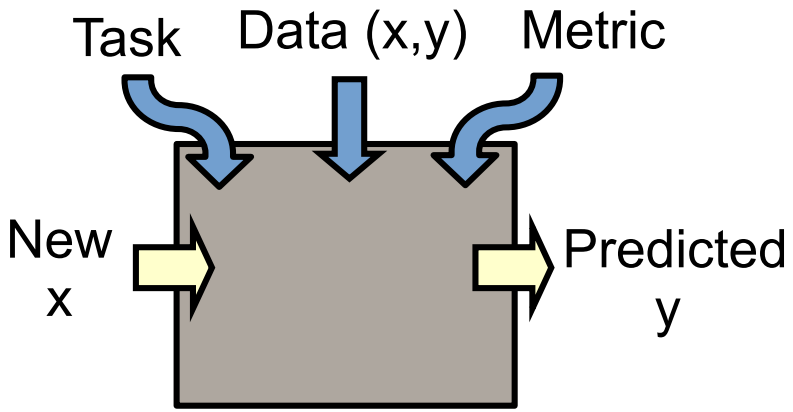
Lecture 11 (Support Vector Machines)

Spring 2021

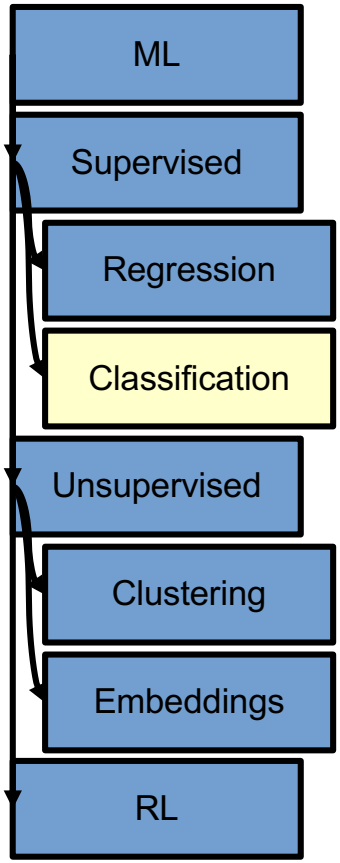
Finale Doshi-Velez and David C. Parkes
Harvard Computer Science

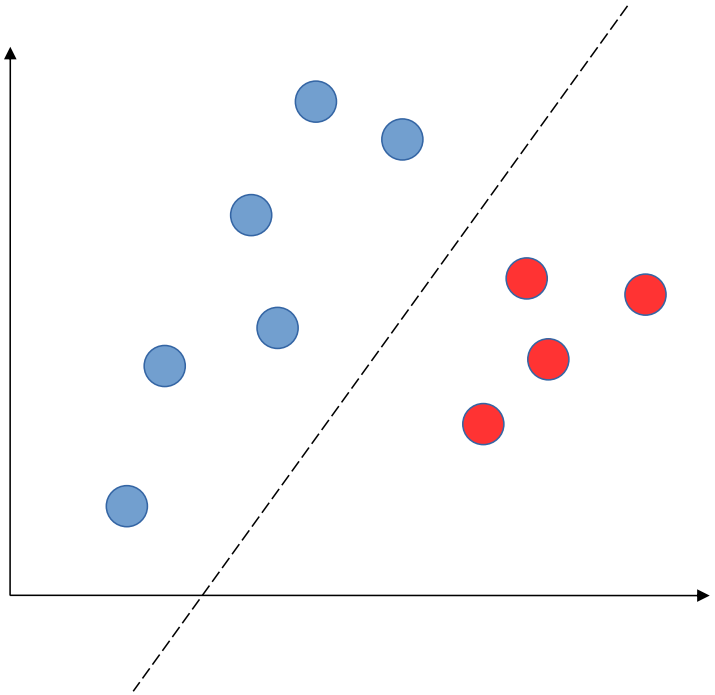
Machine Learning Taxonomy

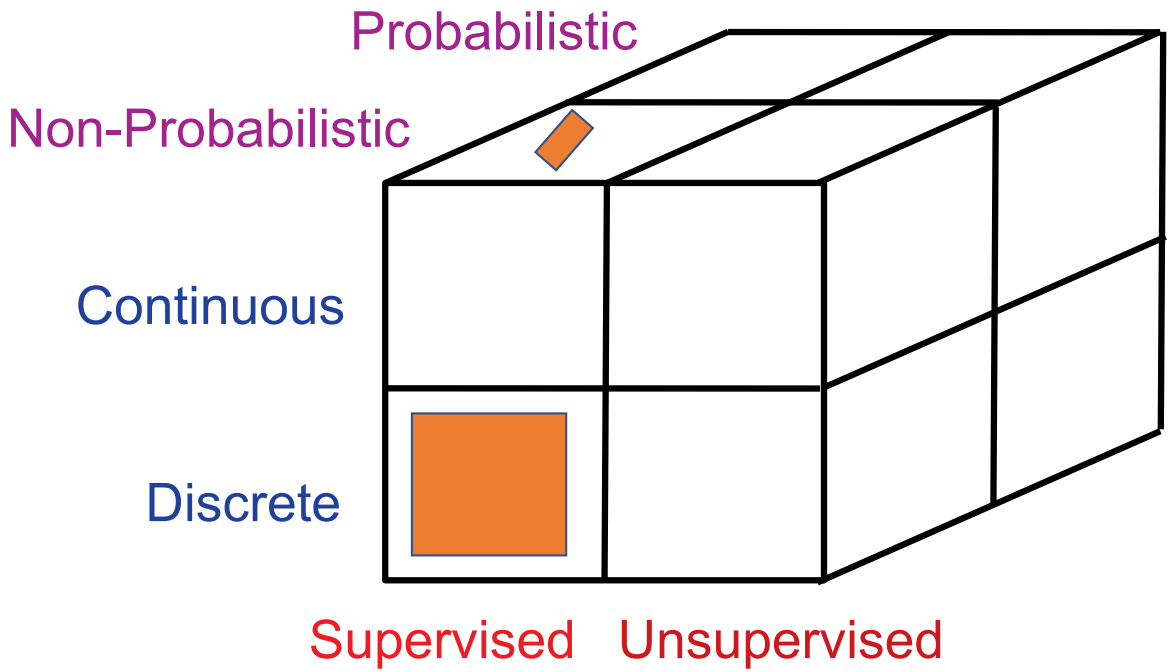




Terminology: Classification



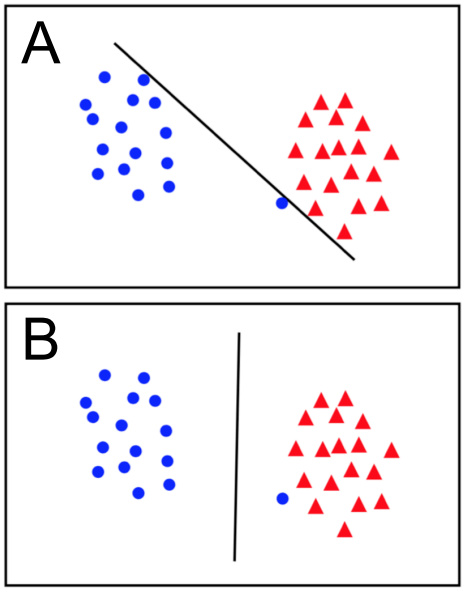




+ graphical models, reinforcement learning

Today: Support Vector Machines

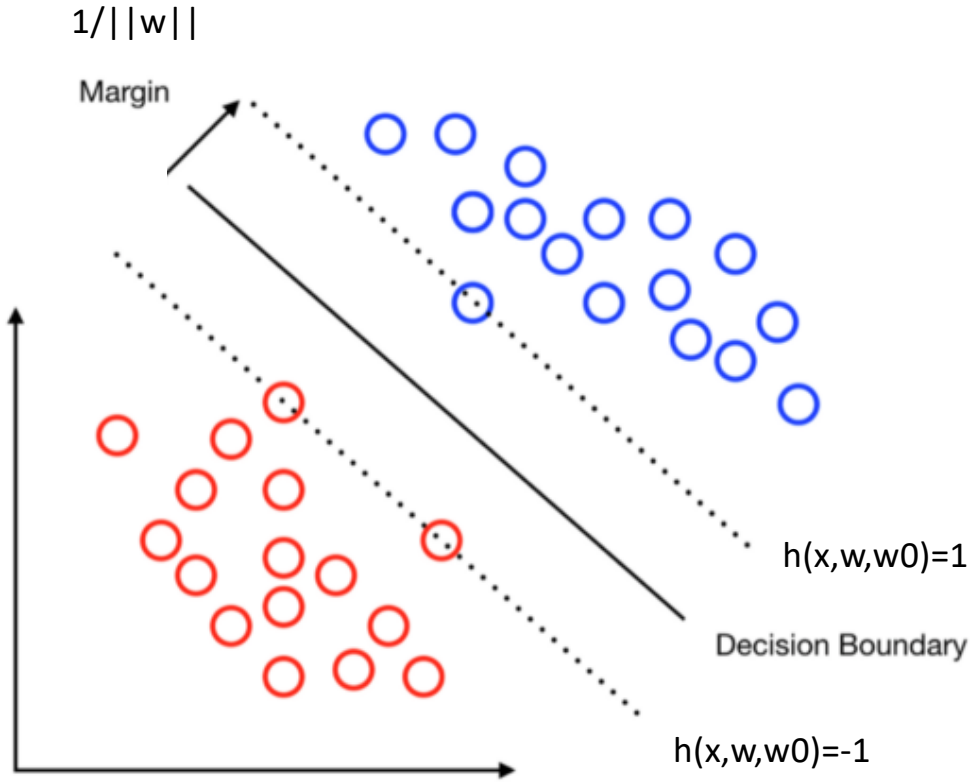
- Last class: **Max-margin methods**, led to "hard margin" and "soft margin" formulations, convex training problems, and coherent theory
- Today: work in the dual formulation, get to SVMs
  - "support vectors"
  - the "Kernel trick", which handles basis functions very nicely
- SVMs have a record of very successful application, for example:
  - Predicting a cancer type from cell samples
  - Fake news characterization
  - Predicting the function of proteins



May prefer B over A

Max-margin Methods

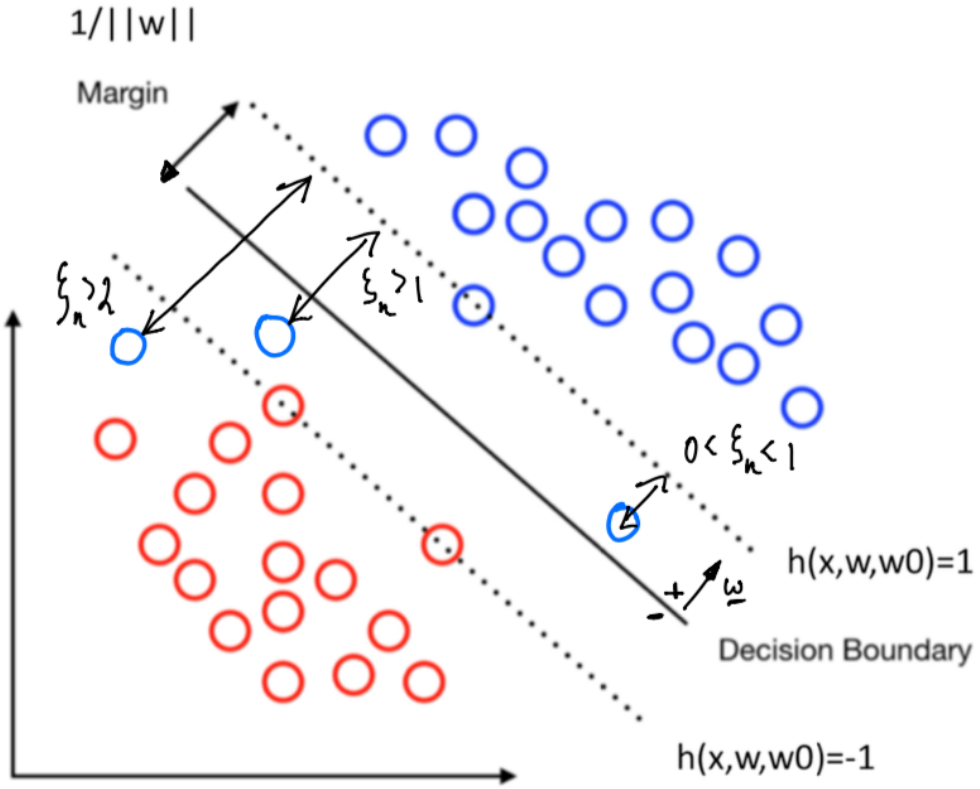
https://medium.com/analytics-vidhya/support-vector-machines-with-amazon-food-reviews-9fe0428e09ef



Hard margin: Find separator to max the min distance(normalized, unsigned, orthogonal) to decision boundary (the “margin”)

Max-margin Methods

https://medium.com/analytics-vidhya/support-vector-machines-with-amazon-food-reviews-9fe0428e09ef



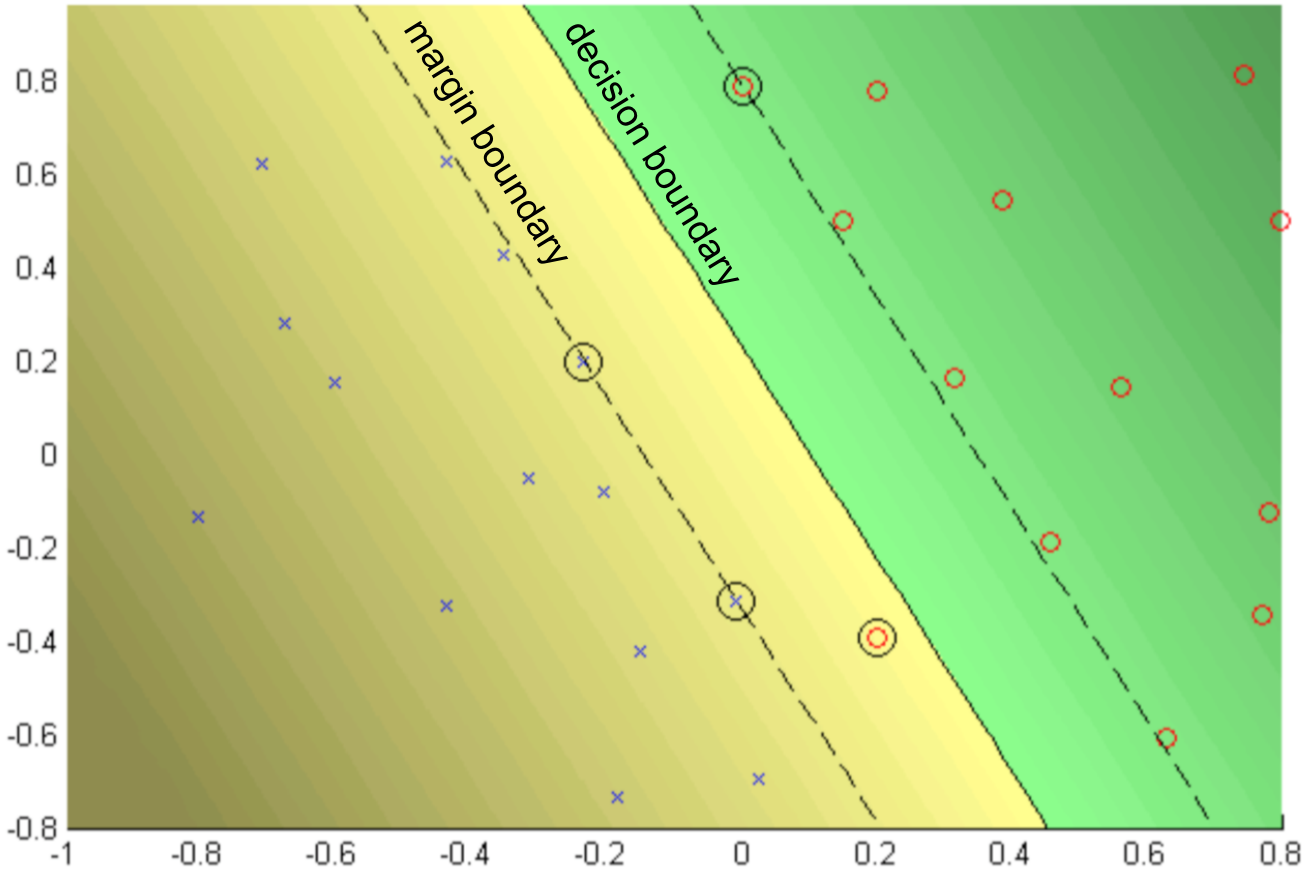
Hard margin: Find separator to max the min distance(normalized, unsigned, orthogonal) to decision boundary (the “margin”)

Soft margin: Allow some examples inside the margin, or misclassified, tradeoff this degree of error with the margin on other examples (larger C, closer to hard margin, less regularization)

$$\min_{\mathbf{w},w_0,\boldsymbol{\xi}\geq 0} \quad \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i$$

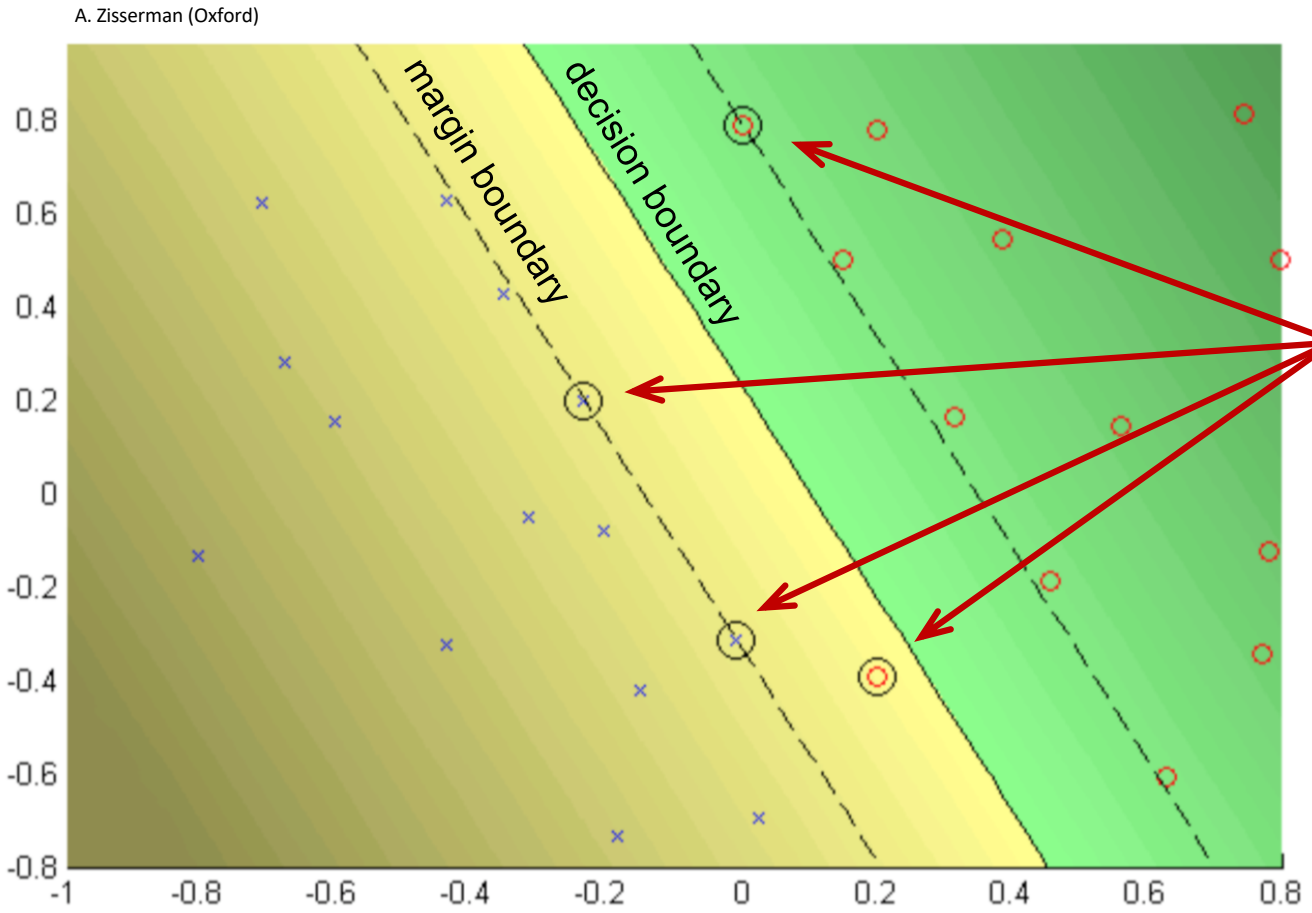
Support Vectors

A. Zisserman (Oxford)



The decision boundary can be described through a weighted vote on “support vectors”

Support Vectors



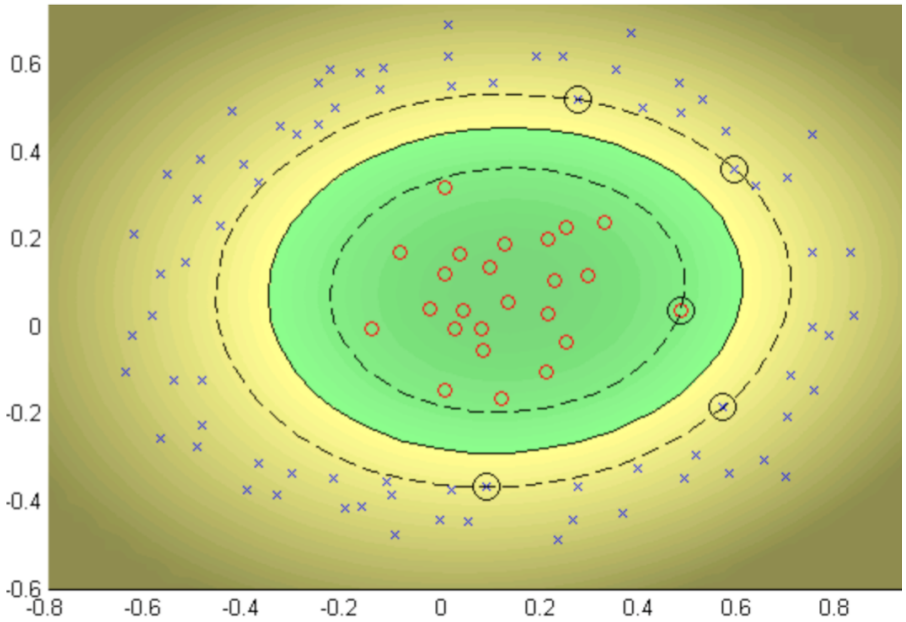
The decision boundary can be described through a weighted vote on “support vectors”

Here there are four support vectors (circled); three on the margin boundaries, one inside the margin region

Can be inside margin region for soft-margin

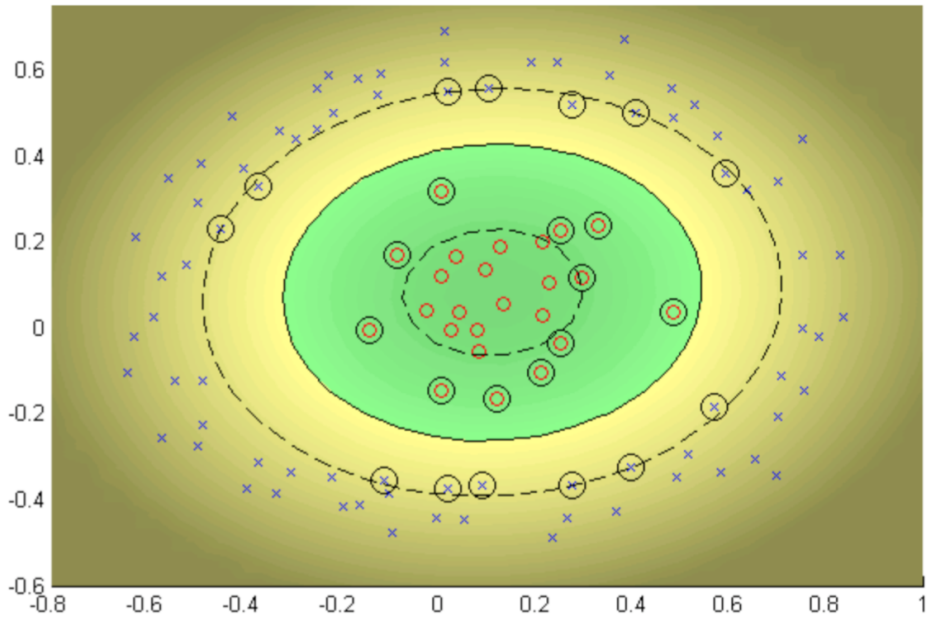
Non-linear basis function

A. Zisserman (Oxford)



Gaussian kernel, C=infinity
5 support vectors (circled)

C = infinity, and so hard margin.



Gaussian kernel, C=10
24 support vectors (circled)

(C smaller, regularization, some examples inside the margin, “bending less” to data)