CS688: Web-Scale Image Retrieval

Basic Classification and Learning Methods

Sung-Eui Yoon
(윤성의)

Course URL:
http://sglab.kaist.ac.kr/~sungeui/IR

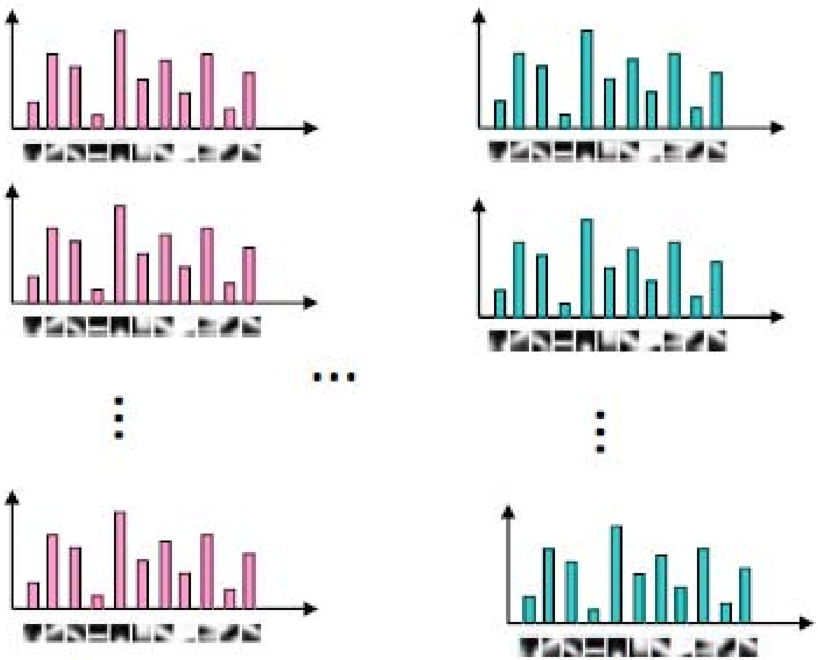
KAIST

Class Objectives

- **Data driven techniques**
  - **Nearest neighbor classifiers**
- **Basic learning methods**
  - **Support Vector Machine (SVM)**

Discriminative classifiers

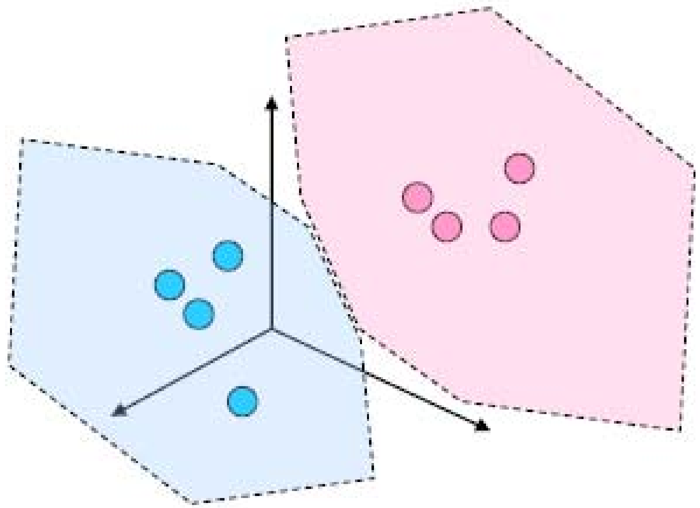
category models



Class 1

Class N

Model space

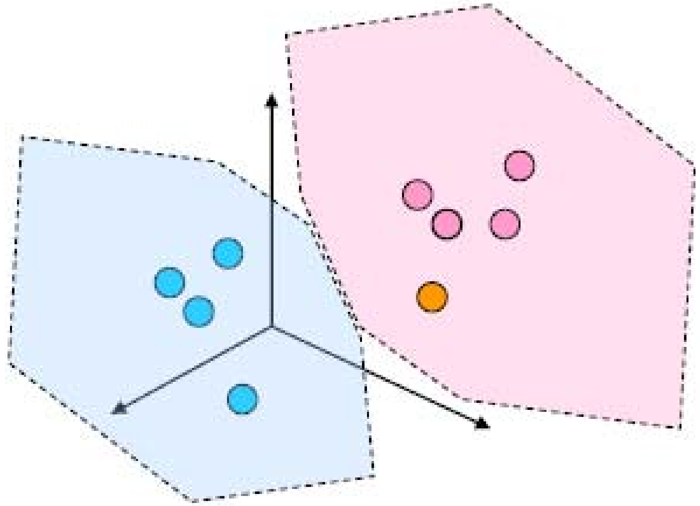


Discriminative classifiers

Query image

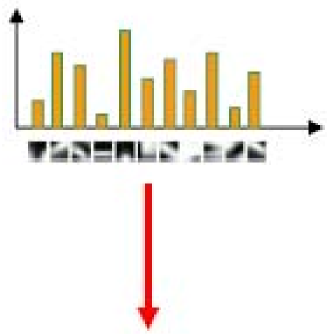
Winning class: pink

Model space



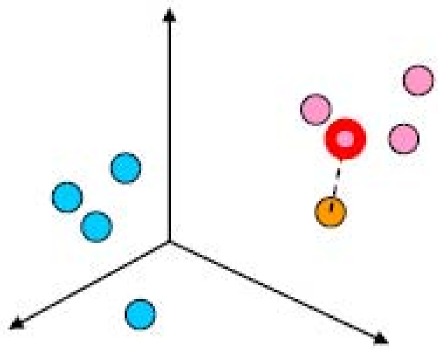
Nearest Neighbors classifier

Query image



Winning class: pink

Model space

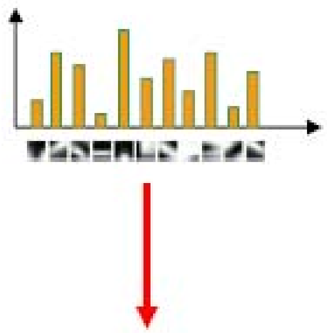


- Assign label of nearest training data point to each test data point

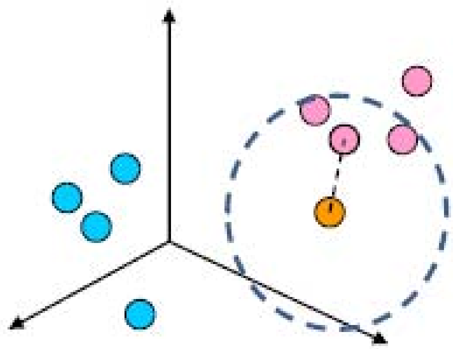
K- Nearest Neighbors classifier

Query image

Model space



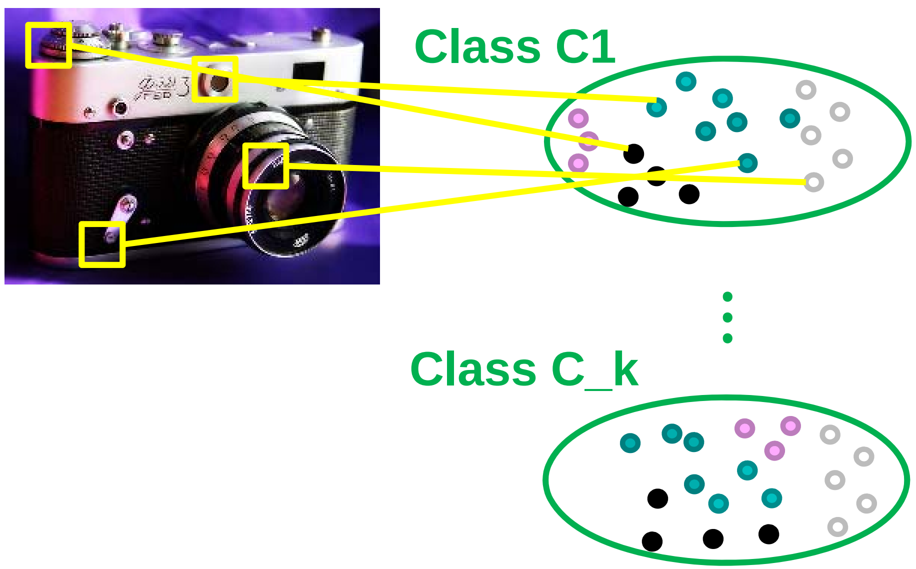
Winning class: pink



- For a new point, find the k closest points from training data
- Labels of the k points "vote" to classify
- Works well provided there is lots of data and the distance function is good

Naïve Bayes Nearest Neighbor (NBNN) Classifier [CVPR 08]

- Extract and collect features for each category

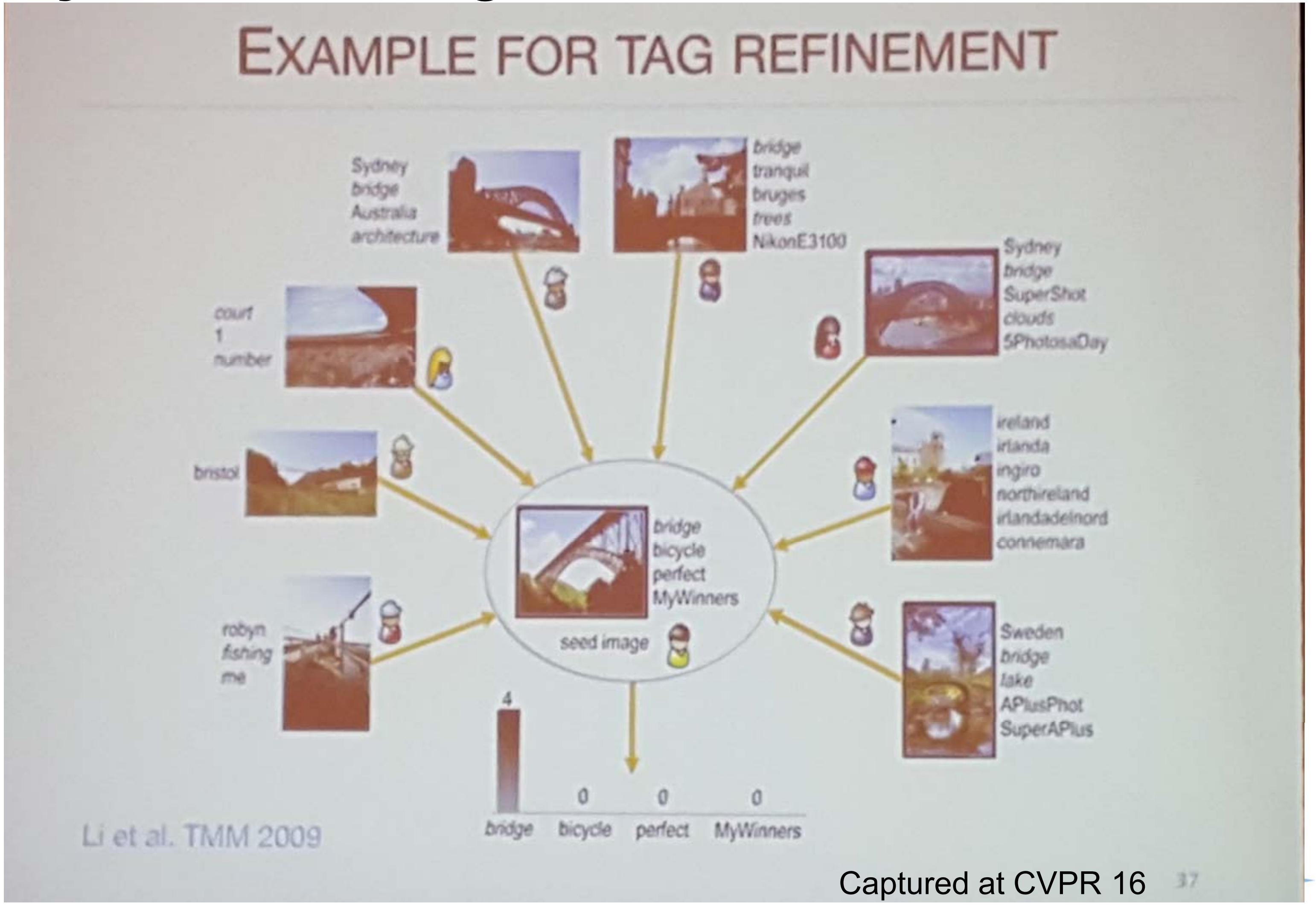


At runt time:

- Extract features for a query image
- Measure their distances from all the categories
- Pick the category w/ the lowest distance

KNN for Tag Transfer

- **Identify similar images and transfer their tags**



Hashing techniques

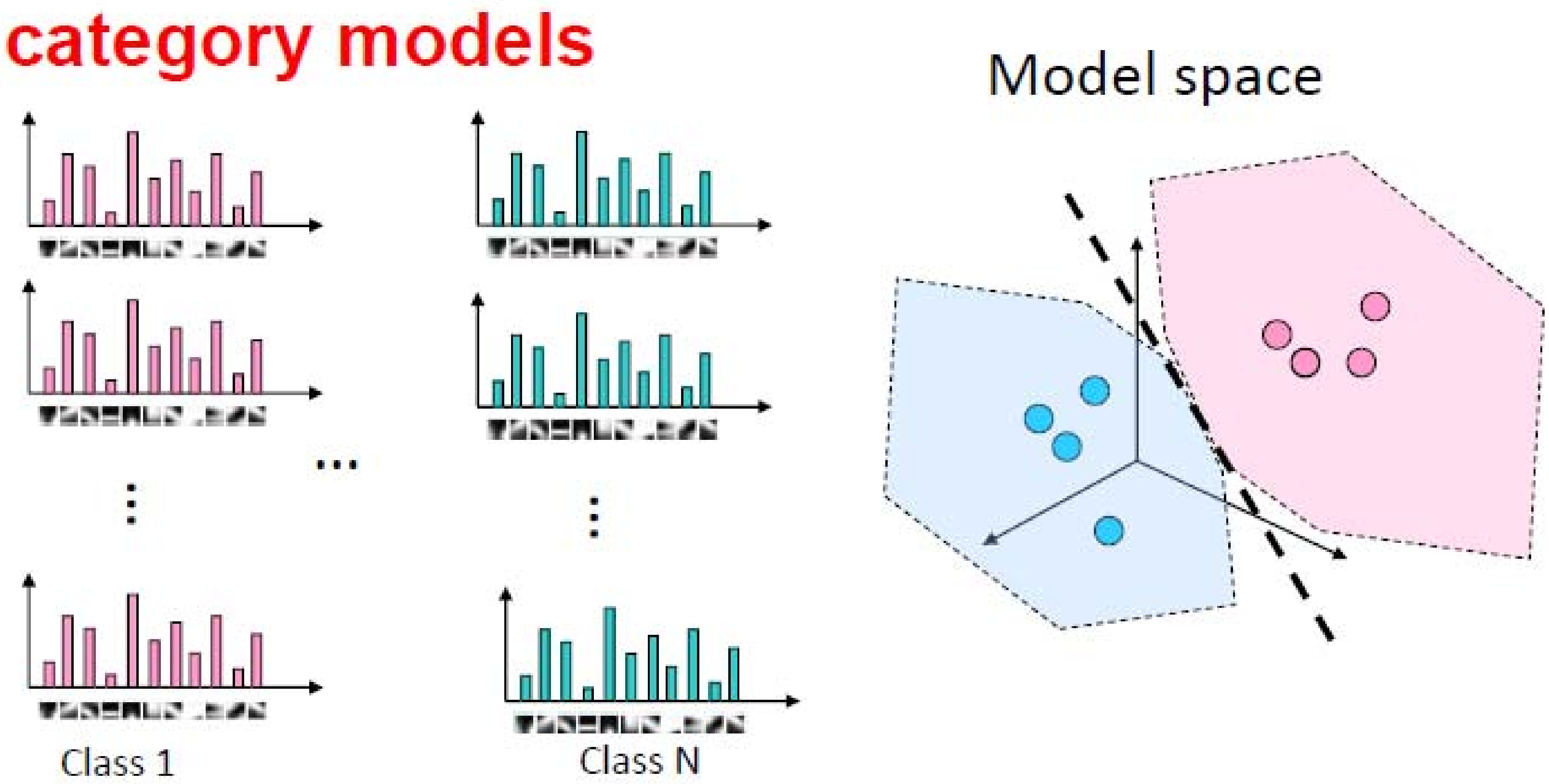
- Fast in high-dimensional problems
  - E.g., Locality sensitive hashing
- Used for binary code embedding to compute compact representation
  - Will be discussed later

Class Objectives

- Data driven techniques
  - Nearest neighbor classifiers

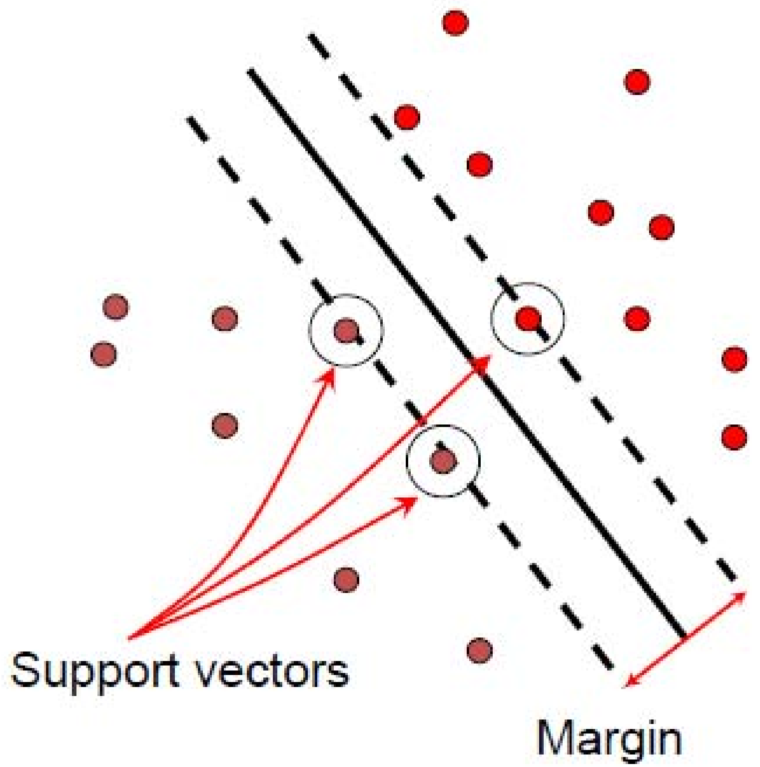
- Basic learning methods
  - Support Vector Machine (SVM)

Discriminative classifiers (linear classifier)



Support vector machines

- Find hyperplane that maximizes the *margin* between the positive and negative examples



Support vectors: $\mathbf{x}_i \cdot \mathbf{w} + b = \pm 1$

Distance between point and hyperplane: $\frac{|\mathbf{x}_i \cdot \mathbf{w} + b|}{||\mathbf{w}||}$

Margin = $2 / ||\mathbf{w}||$

Solution: $\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$

Classification function (decision boundary):

$$\mathbf{w} \cdot \mathbf{x} + b = \sum_i \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b$$

Credit slide: S. Lazebnik

KAIST

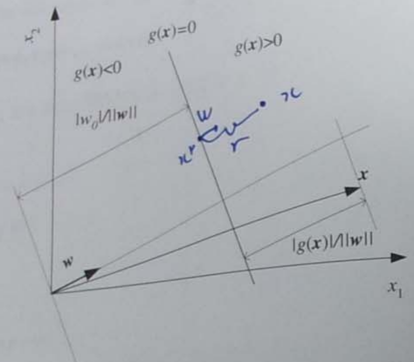


Figure 10.2 The geometric interpretation of the linear discriminant.

where x_p is the normal projection of x onto the hyperplane and r gives us the distance from x to the hyperplane, negative if x is on the negative side, and positive if x is on the positive side (see figure 10.2). Calculating $g(x)$ and noting that $g(x_p) = 0$, we have

$$r = \frac{g(x)}{\|w\|} \tag{10.4}$$

We see then that the distance to origin is

$$r_0 = \frac{w_0}{\|w\|} \tag{10.5}$$

Thus w_0 determines the location of the hyperplane with respect to the origin, and w determines its orientation.

10.3.2 Multiple Classes

When there are $K > 2$ classes, there are K discriminant functions. When they are linear, we have

$$g_i(x|w_i, w_{i0}) = w_i^T x + w_{i0} \tag{10.6}$$

$$g(x) = w_1 \cdot x_1 + w_2 \cdot x_2 + w_0 = w^T \cdot x + w_0$$

$$x = x_p + \frac{w}{\|w\|} \cdot r$$

$$\frac{g(x)}{\|w\|} = \frac{w^T}{\|w\|} \cdot x + \frac{w_0}{\|w\|}$$

$$\frac{g(x_p)}{\|w\|} = 0 = \frac{w^T}{\|w\|} \cdot x_p + \frac{w_0}{\|w\|}$$

$$\Rightarrow \frac{w_0}{\|w\|} = -\frac{w^T}{\|w\|} x_p$$

$$\frac{g(x)}{\|w\|} = \frac{w^T}{\|w\|} \cdot x - \frac{w^T}{\|w\|} \cdot x_p = \frac{w^T}{\|w\|}(x - x_p) = 1 \cdot r \cdot \cos(0^\circ) = r.$$

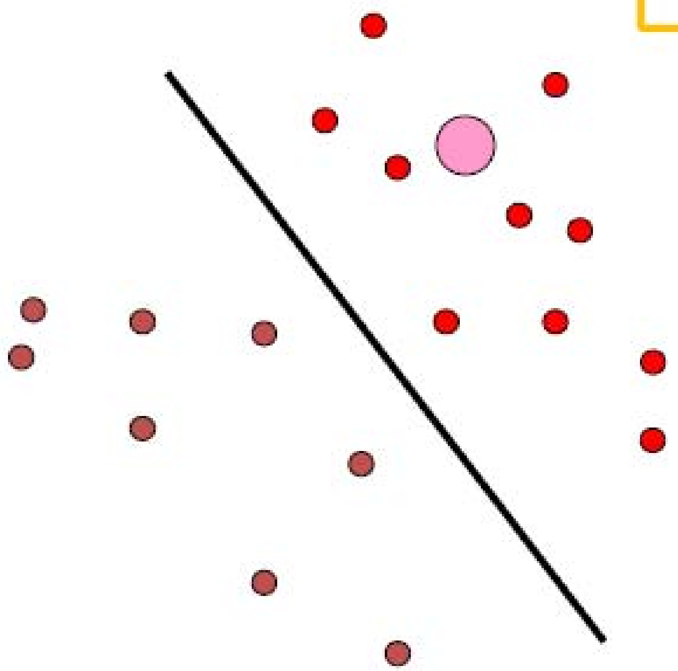
기록자 서명(Signature) 일자(Date)

Support vector machines

- Classification

$$\mathbf{w} \cdot \mathbf{x} + b = \sum_i \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b$$

Test point



Margin

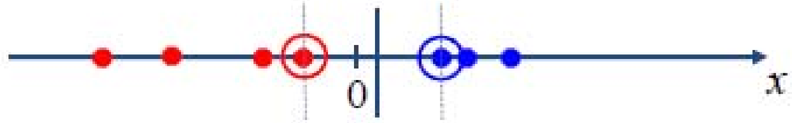
$$if \quad \mathbf{x} \cdot \mathbf{w} + b \geq 0 \rightarrow \quad class\ 1$$

$$if \quad \mathbf{x} \cdot \mathbf{w} + b < 0 \rightarrow \quad class\ 2$$

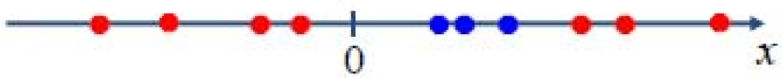
C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, 1998

Nonlinear SVMs

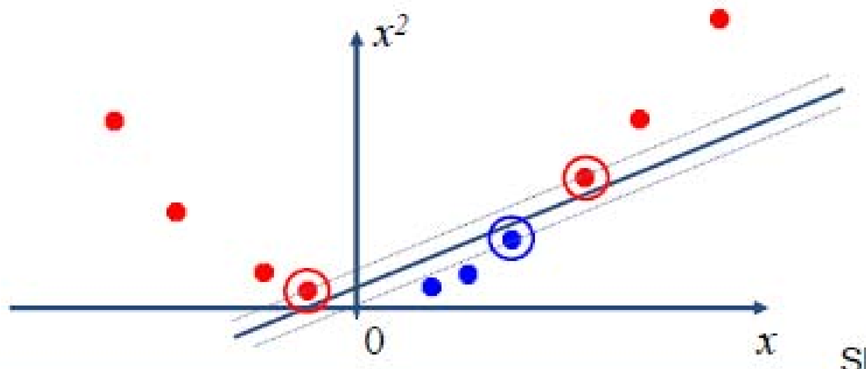
- Datasets that are linearly separable work out great:



- But what if the dataset is just too hard?



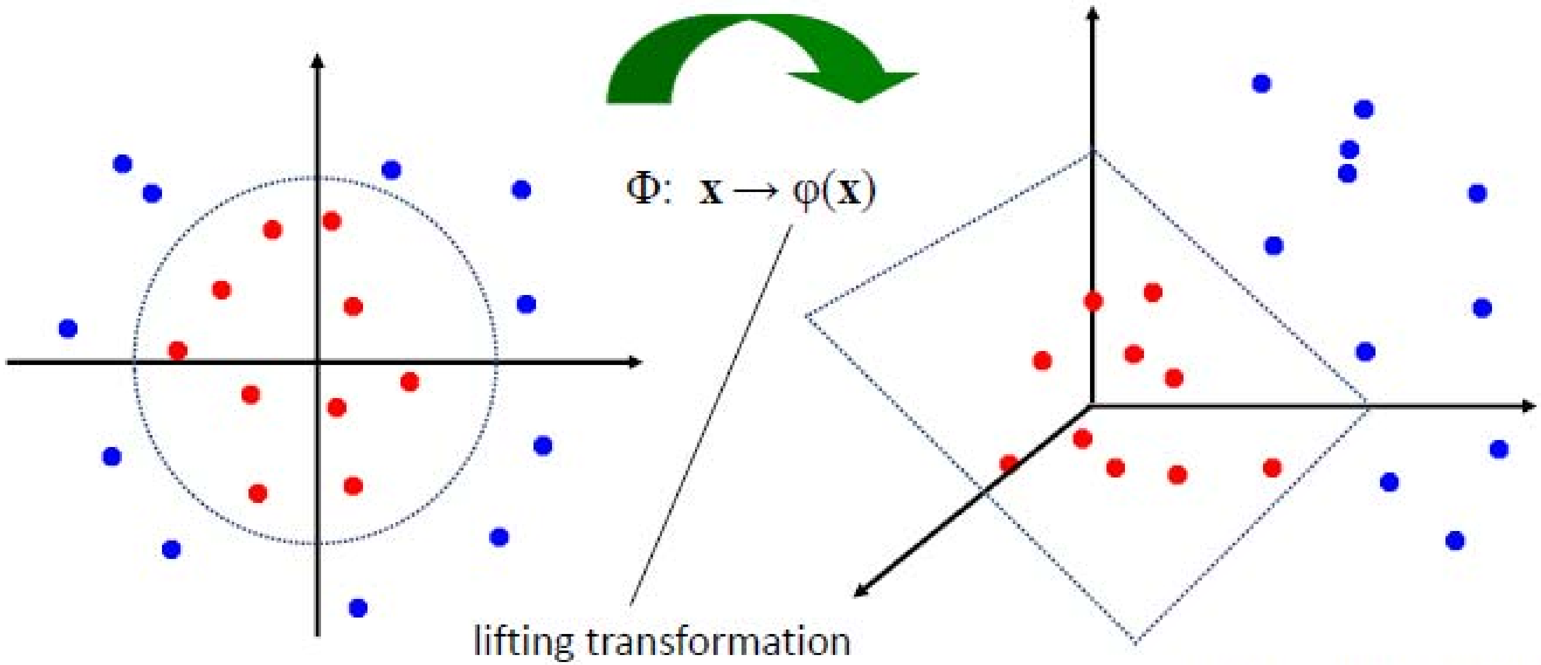
- We can map it to a higher-dimensional space:



Slide credit: Andrew Moore

Nonlinear SVMs

- General idea: the original input space can always be mapped to some higher-dimensional feature space where the training set is separable:



Slide credit: Andrew Moore

What about multi-class SVMs?

- No "definitive" multi-class SVM formulation
- In practice, we have to obtain a multi-class SVM by combining multiple two-class SVMs
- One vs. others
  - Traning: learn an SVM for each class vs. the others
  - Testing: apply each SVM to test example and assign to it the class of the SVM that returns the highest decision value
- One vs. one
  - Training: learn an SVM for each pair of classes
  - Testing: each learned SVM "votes" for a class to assign to the test example

Credit slide: S. Lazebnik

SVMs: Pros and cons

- Pros
  - Many publicly available SVM packages: http://www.kernel-machines.org/software
  - Kernel-based framework is very powerful, flexible
  - SVMs work very well in practice, even with very small training sample sizes

- Cons
  - No "direct" multi-class SVM, must combine two-class SVMs
  - Computation, memory
    - During training time, must compute matrix of kernel values for every pair of examples
    - Learning can take a very long time for large-scale problems

Class Objectives were:

- **Data driven techniques**
  - **Nearest neighbor classifiers**

- **Basic learning methods**
  - **Support Vector Machine (SVM)**

Homework for Every Class

- **Go over the next lecture slides**
- **Come up with one question on what we have discussed today**
  - **1 for typical questions (that were answered in the class)**
  - **2 for questions with thoughts or that surprised me**

- **Write questions at least 4 times**

KAIST

Next Time…

- Deep learning