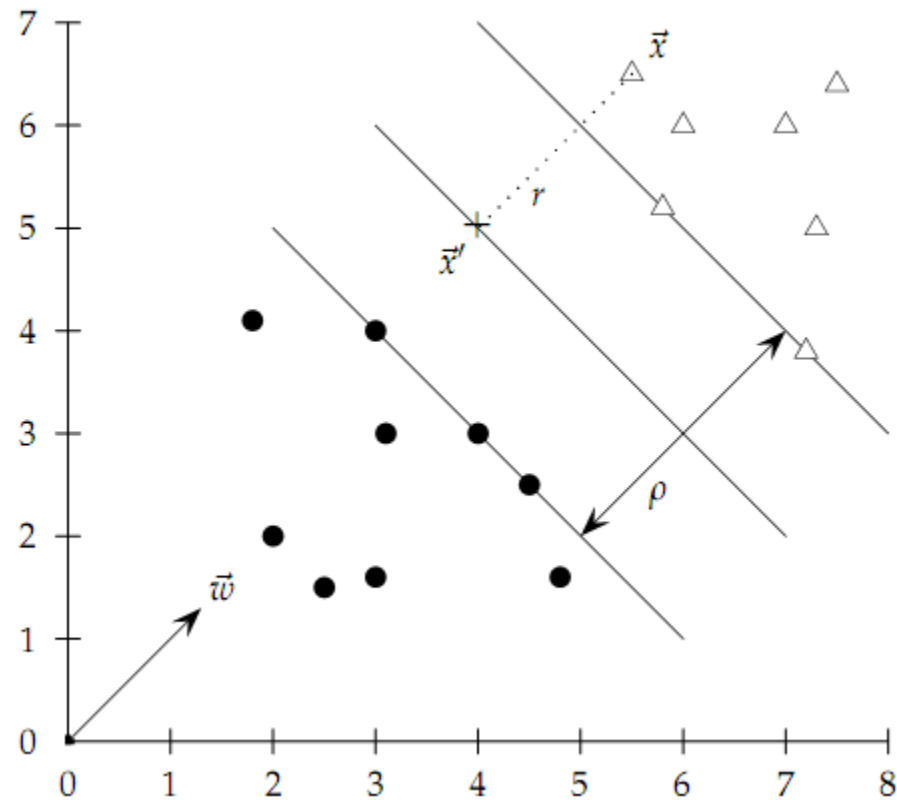


# Support vector machines and machine learning on documents

By cerror

# Support vector machines



# Linear classifier

$$f(\vec{x}) = \text{sign}(\vec{w}^T \vec{x} + b)$$

$$f(\vec{x}) = (\vec{w}^T \vec{x} + b)$$

# Quadratic optimization problems

- Minimize (in  $w, b$ )
  - $\frac{1}{2} \|w\|^2$
- subject to (for any  $i = 1, \dots, n$ )
  - $y_i(w \cdot x_i - b) \geq 1$

# Quadratic optimization

$$\text{Minimize } f(\mathbf{x}) = \mathbf{c}\mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x}$$

subject to  $\mathbf{A}\mathbf{x} \geq \mathbf{b}$  and  $\mathbf{x} \geq \mathbf{0}$

# Dual form

Find  $\alpha_1, \dots, \alpha_N$  such that  $\sum \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \vec{x}_i^T \vec{x}_j$  is maximized, and

- $\sum_i \alpha_i y_i = 0$
- $\alpha_i \geq 0$  for all  $1 \leq i \leq N$

$$\vec{w} = \sum \alpha_i y_i \vec{x}_i$$

$$b = y_k - \vec{w}^T \vec{x}_k \text{ for any } \vec{x}_k \text{ such that } \alpha_k \neq 0$$

# Extensions to the SVM model

- Soft margin classification
- Multiclass SVMs
- Nonlinear SVMs

# Soft margin classification

Find  $\vec{w}$ ,  $b$ , and  $\xi_i \geq 0$  such that:

- $\frac{1}{2}\vec{w}^T\vec{w} + C \sum_i \xi_i$  is minimized
- and for all  $\{(\vec{x}_i, y_i)\}$ ,  $y_i(\vec{w}^T\vec{x}_i + b) \geq 1 - \xi_i$

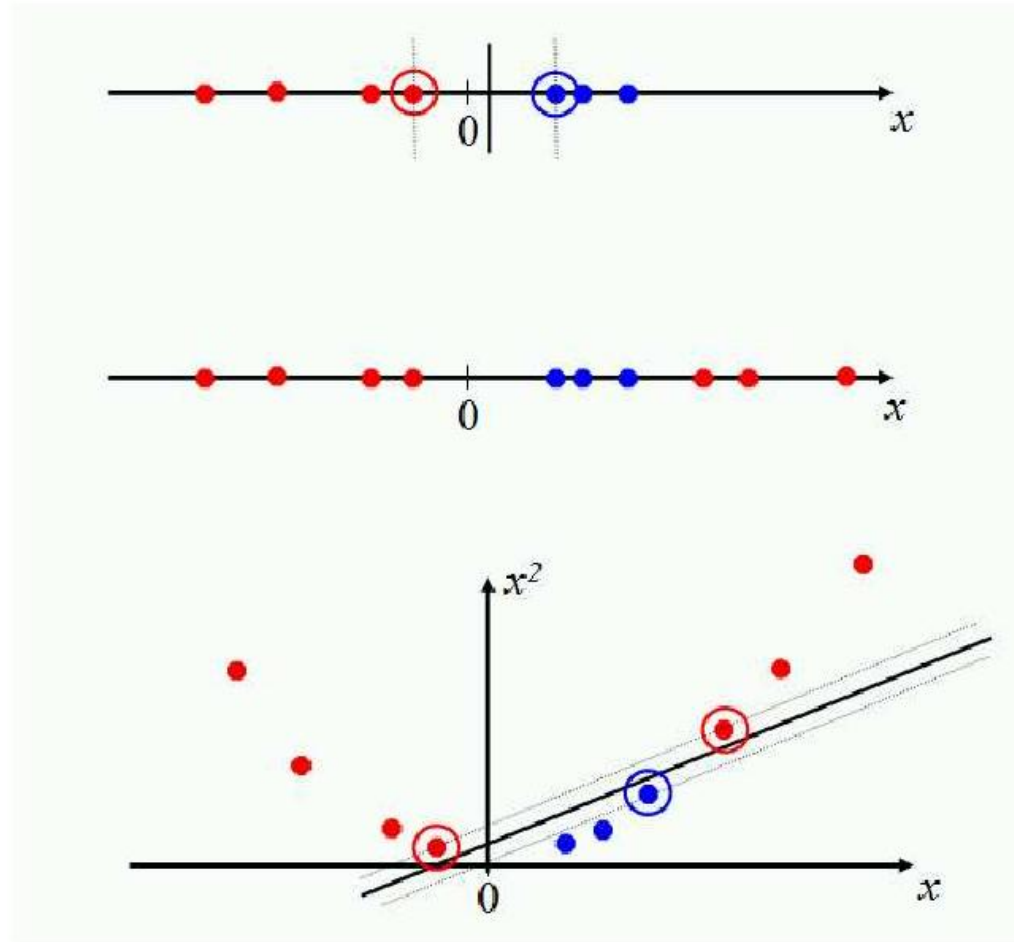
Find  $\alpha_1, \dots, \alpha_N$  such that  $\sum \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \vec{x}_i^T \vec{x}_j$  is maximized, and

- $\sum_i \alpha_i y_i = 0$
- $0 \leq \alpha_i \leq C$  for all  $1 \leq i \leq N$

# Multiclass SVMs

- any-of classification
- one-of classification
- $y = \operatorname{argmax}(w' \Phi(x, y'))$

# Nonlinear SVMs



# Nonlinear SVMs

- Nonlinear Kernels

# Issues in the classification of text documents

- Choosing what kind of classifier to use
- Improving classifier performance

# Choosing what kind of classifier to use

- How much training data is there currently available?
  - no labeled training data: hand-written rules
  - semi-supervised training methods
  - reasonable amount of labeled data
  - a huge amount of data are available

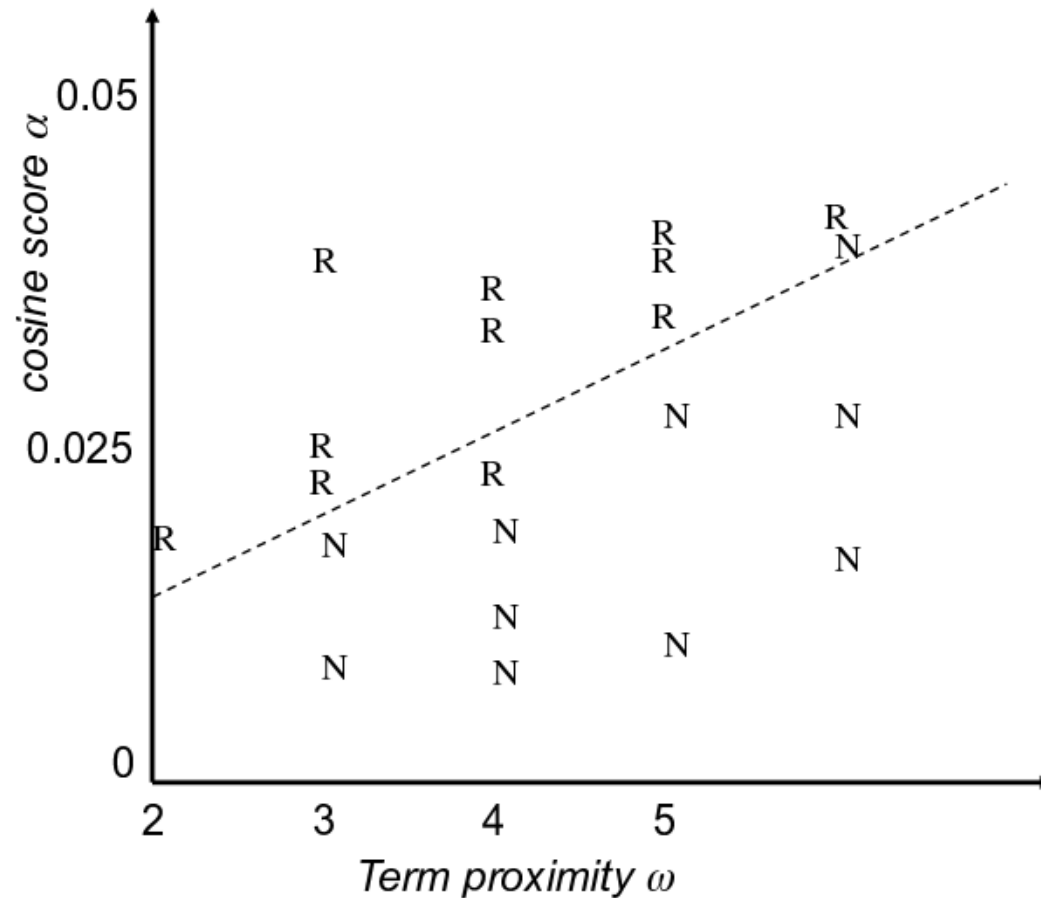
# Improving classifier performance

- Large and difficult category taxonomies
- Features for text
- Document zones in text classification
  - Upweighting document zones
  - Separate feature spaces for document zones
  - Connections to text summarization

# Machine learning methods in ad hoc information retrieval

- A simple example of machine-learned scoring
- Result ranking by machine learning

# A simple example of machine-learned scoring



# Result ranking by machine learning

$$\Phi(d_i, d_j, q) = \psi(d_i, q) - \psi(d_j, q)$$

$$\vec{w}^T \Phi(d_i, d_j, q) > 0 \quad \text{iff} \quad d_i \prec d_j$$

Find  $\vec{w}$ , and  $\xi_{i,j} \geq 0$  such that:

- $\frac{1}{2} \vec{w}^T \vec{w} + C \sum_{i,j} \xi_{i,j}$  is minimized
- and for all  $\{\Phi(d_i, d_j, q) : d_i \prec d_j\}$ ,  $\vec{w}^T \Phi(d_i, d_j, q) \geq 1 - \xi_{i,j}$