CSC 411: Lecture 15: Support Vector Machine

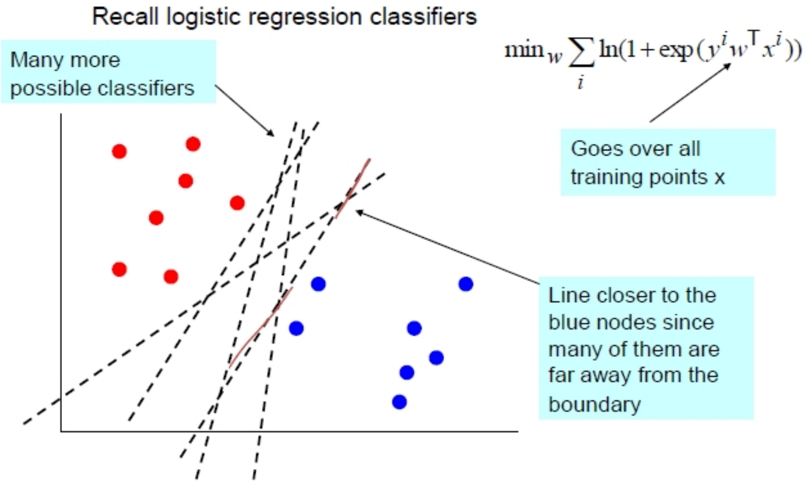
Raquel Urtasun & Rich Zemel

University of Toronto

Nov 11, 2015

Today

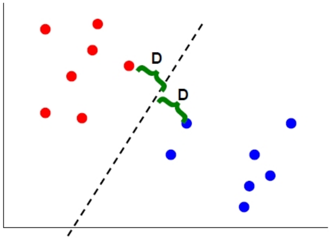
- Margin
- Max-margin classification



$$y = \begin{cases} 1 & \text{if } (\mathbf{w}^T\mathbf{x} + b) \geq 0 \\ -1 & \text{if } (\mathbf{w}^T\mathbf{x} + b) < 0 \end{cases}$$

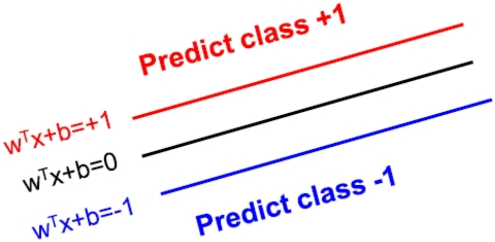
Max margin classification

- Instead of fitting all the points, focus on boundary points
- Aim: learn a boundary that leads to the largest margin (buffer) from points on both sides



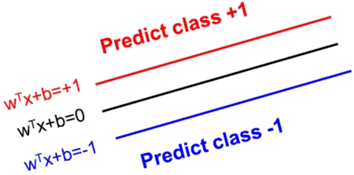
- Why: intuition; theoretical support; and works well in practice
- Subset of vectors that support (determine boundary) are called the support vectors

- Max margin classifier: inputs in margin are of unknown class



$$y = \begin{cases} 1 & \text{if } \mathbf{w}^T\mathbf{x} + b \geq 1 \\ -1 & \text{if } \mathbf{w}^T\mathbf{x} + b \leq -1 \\ \text{Undefined} & \text{if } -1 \leq \mathbf{w}^T\mathbf{x} + b \leq 1 \end{cases}$$

Geometry of the Problem



- The vector **w** is orthogonal to the $+1$ plane.
  If **u** and **v** are two points on that plane, then

$$\mathbf{w}^T(\mathbf{u} - \mathbf{v}) = 0$$

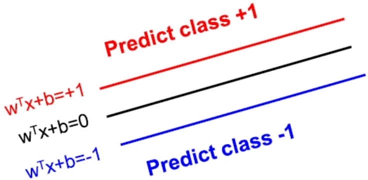
- Same is true for -1 plane
- Also: for point $\mathbf{x}_+$ on $+1$ plane and $\mathbf{x}_-$ nearest point on -1 plane:

$$\mathbf{x}_+ = \lambda \mathbf{w} + \mathbf{x}_-$$

Computing the Margin

- Also: for point $\mathbf{x}_+$ on $+1$ plane and $\mathbf{x}_-$ nearest point on -1 plane:

$$\mathbf{x}_+ = \lambda \mathbf{w} + \mathbf{x}_-$$



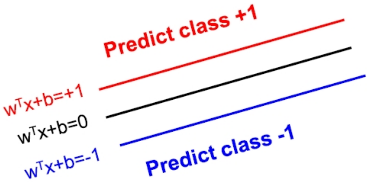
$$\begin{aligned} \mathbf{w}^T\mathbf{x}_+ + b &= 1 \\ \mathbf{w}^T(\lambda\mathbf{w} + \mathbf{x}_-) + b &= 1 \\ \mathbf{w}^T\mathbf{x}_- + b + \lambda\mathbf{w}^T\mathbf{w} &= 1 \\ -1 + \lambda\mathbf{w}^T\mathbf{w} &= 1 \end{aligned}$$

Therefore

$$\lambda = \frac{2}{\mathbf{w}^T\mathbf{w}}$$

Computing the Margin

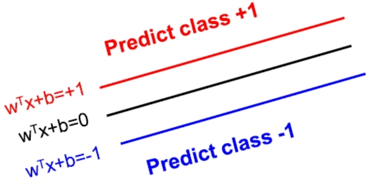
- Define the margin M to be the distance between the $+1$ and -1 planes
- We can now express this in terms of $\mathbf{w}$ to maximize the margin we minimize the length of $\mathbf{w}$



$$
\begin{aligned}
M &= ||\mathbf{x}_+ - \mathbf{x}_-|| \\
&= ||\lambda \mathbf{w}|| = \lambda \sqrt{\mathbf{w}^T\mathbf{w}} \\
&= 2\frac{\sqrt{\mathbf{w}^T\mathbf{w}}}{\mathbf{w}^T\mathbf{w}} = \frac{2}{\sqrt{\mathbf{w}^T\mathbf{w}}} = \frac{2}{||\mathbf{w}||}
\end{aligned}
$$

Learning a Margin-Based Classifier

- We can search for the optimal parameters ($\mathbf{w}$ and b) by finding a solution that:
  1. Correctly classifies the training examples: $\{(\mathbf{x}^{(i)}, t^{(i)})\}_{i=1}^{N}$
  2. Maximizes the margin (same as minimizing $\mathbf{w}^T\mathbf{w}$)



$$\min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2$$

$$s.t. \forall i \quad (\mathbf{w}^T\mathbf{x}^{(i)} + b)t^{(i)} \geq 1,$$

- This is call the primal formulation of Support Vector Machine (SVM)
- Can optimize via projective gradient descent, etc.
- Apply Lagrange multipliers: formulate equivalent problem