Home Page

Title Page

Contents

Page 1 of 41

Go Back

Full Screen

Close

Quit

Support Vector Machines

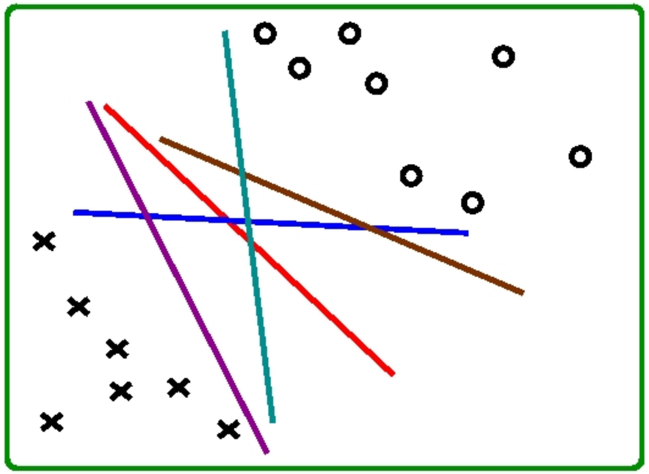
Sumantra Dutta Roy
Department of Electrical Engineering
Indian Institute of Technology Delhi
Hauz Khas, New Delhi - 110 016, INDIA.
http://www.cse.iitd.ac.in/∼sumantra
sumantra@ee.iitd.ac.in

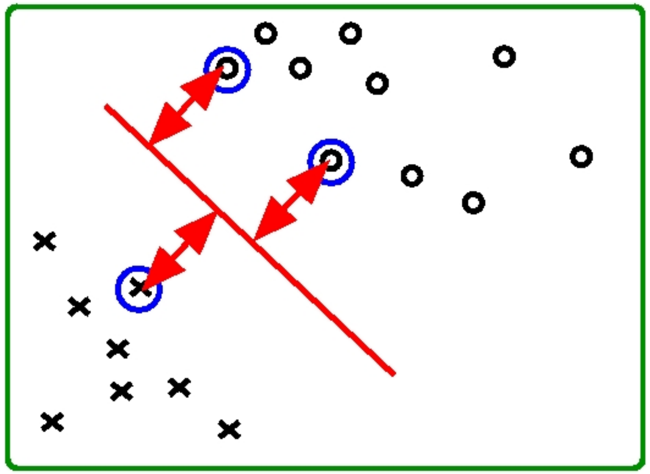
Introduction

- SVMs: sparse solutions: support vectors
- Convex optimisation: local optimal is also global!
- Philosophy: different classifier formulations seek to optimise different criteria, make different assumptions, each fine in its own right
- Fisher: maximise 1-D projected margins
- SVM: maximise margins
- 2-class: generalisable to K classes: 1-1 or 1-rest
- 2 classes: ± 1 for notational convenience, not $0/1$

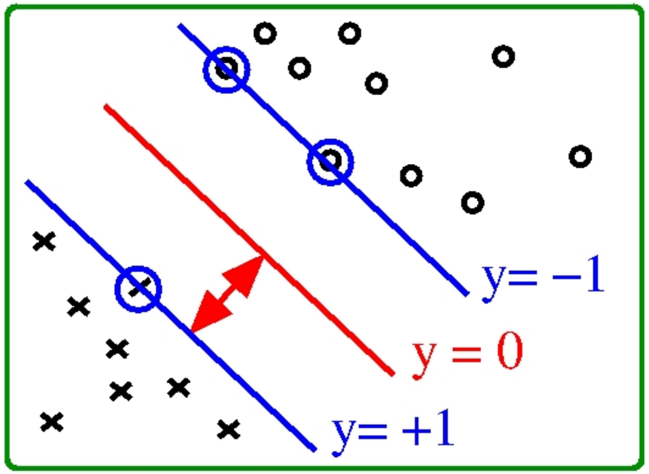
Maximum Margin Classifiers

- 2-class restricted possible non-linearity:
- $y(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}) + b$. $y(\mathbf{x})$ is the model for target t. b: bias, $\boldsymbol{\phi}(\cdot)$: feature space transformation
- training $\{\mathbf{x}_1, \ldots \mathbf{x}_N\}$ targets $\{t_1, \ldots t_N\}$, $t_i \in \{-1, +1\}$
- New data points $\mathbf{x}$ classified acc to $y(\mathbf{x})$'s sign
- Assume that the training data is linearly separable in the feature space i.e., $\exists$ at least one $\mathbf{w}$ & b:
- $y(\mathbf{x}_i) < 0$ for $t_i = -1$ and $y(\mathbf{x}_i) > 0$ for $t_i = +1$
- Combined: $t_i\, y(\mathbf{x}_i) > 0\ \forall$ training data $\{\mathbf{x}_i, t_i\}$
- Intuition: if multiple solutions, find the one which will give the smallest generalisation error





- Infinite possibilities
- Choose acc to some optimisation criterion
- min dist of a point from the decision boundary
- $\triangleq$ margin, to maximise
- Implicit form of the eqn of a line $w_2 x_2 + w_1 x_1 + b = 0$
- The two intercept form $\frac{x_2}{(-b/w_2)} + \frac{x_1}{(-b/w_1)} = 1$
- The slope-intercept form $x_2 = (\frac{-w_1}{w_2})x_1 + (\frac{-b}{w_2})$
- Take-home point#1: $\mathbf{w}$ determines the slope
- Take-home point#2: b: scaled distance from the origin. Why? $\frac{b}{||\mathbf{w}||}$ is the distance from the origin.



- Margin: min dist b/w decision boundary & any sample
- symmetric, by defn above
- Don't worry about $y = \pm 1$, *yet*
- Aim: Maximise this margin
- The location of the boundary: determined by a small subset of the data points: Support Vectors
- Decision surface: $y(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}) + b = 0$
- Perp dist of $\mathbf{x}$ from $y(\mathbf{x}) = 0$: is given by $\frac{|y(\mathbf{x})|}{||\mathbf{w}||}$
- We want correct classification i.e., $t_i \ y(\mathbf{x}_i) > 0 \ \forall i$
- $|y(\mathbf{x}_i)| = +y(\mathbf{x}_i)$ when $y(\mathbf{x}_i) > 0 \ (t_i = +1)$
- $|y(\mathbf{x}_i)| = -y(\mathbf{x}_i)$ when $y(\mathbf{x}_i) < 0 \ (t_i = -1)$
- $\implies |y(\mathbf{x})| = t_i \ y(\mathbf{x}_i)$, perp dist = $\frac{t_i \ y(\mathbf{x}_i)}{||\mathbf{w}||}$
- Max margin: $\boxed{\arg\max_{\mathbf{w},b}\{\frac{1}{||\mathbf{w}||}\min_i\{t_i \ [\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) + b]\}\}}$

