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 $0/1$

Home Page

Title Page

Contents

Page 3 of 41

Go Back

Full Screen

Close

Quit

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