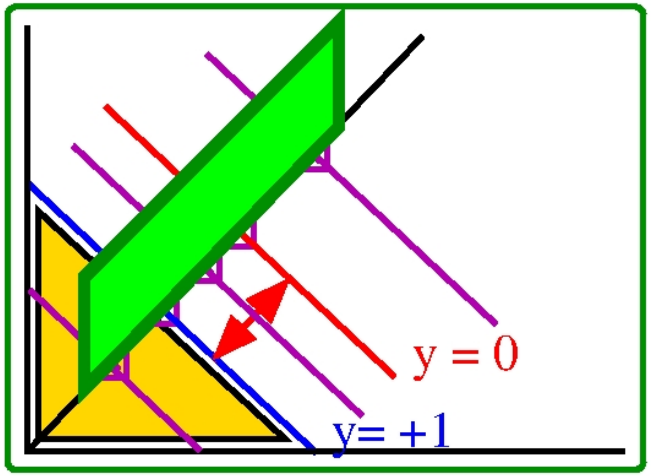

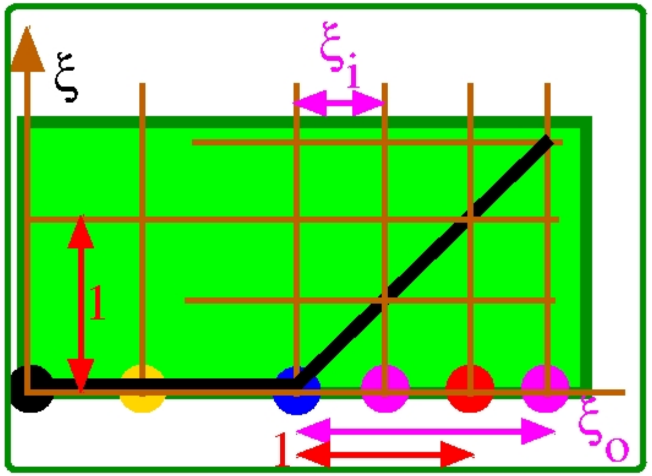

- An outlier excursion point beyond the 'boundary': Consider a point at $\xi_o \triangleq \xi_i$ from 'near' margin.
  - $\frac{\xi_o}{||\mathbf{w}||}$ farther from origin than the 'near' margin
  - eqn: $\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) + (b - 1 + \xi_o) = 0$: $y_i = 1 - \xi_o$
  - $\xi_i = |t_i - y(\mathbf{x}_i)| = |+1 - (1 - \xi_o)| = \xi_i$: $[\xi_o, \xi_o]$
  - Aliter: $\frac{\xi_o - 1}{||\mathbf{w}||}$ farther from the decision boundary
  - eqn: $\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) + b + (\xi_o - 1) = 0$: $y_i = 1 - \xi_o$

Home Page

Title Page

Contents

Page 32 of 41

Go Back

Full Screen

Close

Quit

- Similar formulation for the other side: $y = -1$ class
- $|\cdot|$ imposes symmetry
- For the +1 class, $\xi_i = |t_i - y(\mathbf{x}_i)|$, $\xi_i = 1 - y(\mathbf{x}_i)$
- Required hard region: $y(\mathbf{x}_i) \geq 1$, 'golden' region
- Required (relaxed) region: $y(\mathbf{x}_i) \geq 1 - \xi_i$
- Generalising from the -1 class also, $t_i\, y(\mathbf{x}_i) = 1 - \xi_i$
- Complete (relaxed) region: $t_i\, y(\mathbf{x}_i) \geq 1 - \xi_i$
- Hard to Soft: allow some data be misclassified
- Slack: allow for overlapping class distributions

Home Page

Title Page

Contents

Page 33 of 41

Go Back

Full Screen

Close

Quit

Basic Soft-Margin Formulation

- Framework still sensitive to outliers as misclassification penalty increases linearly with ξ_i
- 1 possible formulation, QP-suitable & num stable
- Primal Formul'n: $\sim$hard-margin primal(simple!)
- To minimise $\frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{N}\xi_i$, 1st term: previous
- 2nd term: penalises far points/excursions
- C: trade-off b/w margin & slack penalty, empirical
- $\sum_{i=1}^{N}\xi_i$: upper bounds # misclassifications Why?
- Correct classification: $\xi_i = 0$, or ≤ 1
- ($\forall$ Miscl: $\xi_i > 1$) $\sum_{i=1}^{K}\xi_i > K \Longrightarrow$ #miscl $< \sum_{i=1}^{K}\xi_i$
- $L(\mathbf{w}, b, \mathbf{a}, \boldsymbol{\xi}, \boldsymbol{\mu}) \triangleq \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} a_i\{t_i\,[\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i) + b] - 1 + \xi_i\} + C\sum_{1=1}^{N}\xi_i - \sum_{i=1}^{N}\mu_i\xi_i$

Home Page

Title Page

Contents

Page 34 of 41

Go Back

Full Screen

Close

Quit

- $L(\mathbf{w}, b, \text{L'mults}) \triangleq$ basic fn - lin combo constrs (≥ 0)
  1. $a_i \geq 0$ (Lagrange)
  2. $t_i y_i \geq 1 - \xi_i \Longrightarrow t_i y_i - 1 + \xi_i \geq 0$ (Golden constr)
  3. $a_i [t_i y_i - 1 + \xi_i] = 0$ (imp) [KKT]
- More? $\xi_i \in [0, 1]$: margin, $\xi_i \geq 1$: outlier
  1. $\xi_i \geq 0$ interpret as a constraint
  2. Have a Lagrange multiplier μ_i: $\mu_i \geq 0$
  3. $\xi_i \mu_i = 0$ (imp) [KKT]
- $L(\mathbf{w}, b, \mathbf{a}, \boldsymbol{\xi}, \boldsymbol{\mu}) \triangleq \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} a_i \{t_i \, [\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) + b] - 1 + \xi_i\} + C \sum_{1=1}^{N} \xi_i - \sum_{i=1}^{N} \mu_i \xi_i$
- Two terms with sum of ξ_is: how?
- 1st: bound on # miscl, other: Lagrange mults
- Constraints? $\sim$ hard margin, $+$ excursions

- $L(\mathbf{w}, b, \mathbf{a}, \boldsymbol{\xi}, \boldsymbol{\mu}) \stackrel{\triangle}{=} \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} a_i \{t_i \, [\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) + b] - 1 + \xi_i\} + C \sum_{1=1}^{N} \xi_i - \sum_{i=1}^{N} \mu_i \xi_i$
- $\{a_i\}$, $\{\xi_i\}$: Lagrange mults. Partial diff wrt $\mathbf{w}$, b, ξ_i
- $\frac{\partial L}{\partial \mathbf{w}^T} = \mathbf{0}$, $\frac{\partial L}{\partial b} = 0$, $\frac{\partial L}{\partial \xi_i} = 0$ 1st 2: same as before!
- $\frac{\partial L}{\partial \mathbf{w}^T} = \frac{1}{2} 2\mathbf{w} - \sum_{i=1}^{N} a_i \, t_i \, \boldsymbol{\phi}(\mathbf{x}_i) = 0$: $\mathbf{w} = \sum_{i=1}^{N} a_i \, t_i \, \boldsymbol{\phi}(\mathbf{x}_i)$
- $\frac{\partial L}{\partial b} = 0$: $\sum_{i=1}^{N} a_i \, t_i = 0$
- $\frac{\partial L}{\partial \xi_i} = 0$: $C - a_i - \mu_i = 0$: $a_i = C - \mu_i$: 'Box Constraints'
- $a_i = C - \mu_i$, Lagrange: $a_i \geq 0$, $\mu_i \geq 0$
- $a_i \geq 0 \implies C - \mu_i \geq 0 \implies \mu_i \leq C$ Add $\mu_i \geq 0$
- $\mu_i \geq 0 \implies C - a_i \geq 0 \implies a_i \leq C$ Add $a_i \geq 0$
- $\boxed{0 \leq \mu_i \leq C}$ $\boxed{0 \leq a_i \leq C}$ a_i, $\mu_i \sim$ dims of a 2-D 'box'