Home Page

Title Page

Contents

Page 34 of 41

Go Back

Full Screen

Close

Quit

- $L(\mathbf{w}, b, \text{L'mults}) \stackrel{\triangle}{=}$ 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]
- $L(\mathbf{w}, b, \mathbf{a}, \boldsymbol{\xi}, \boldsymbol{\mu}) \stackrel{\triangle}{=} \frac{1}{2}||\mathbf{w}||^2 -$
  $\sum_{i=1}^{N} a_i \{t_i \left[\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) + b\right] - 1 + \xi_i\} + C \sum_{1=1}^{N} \xi_i - \sum_{i=1}^{N} \mu_i \xi_i$
- Two terms with sum of ξ_is: how?
- 1st: bound on # miscl, other: Lagrange mults
- Constraints? $\sim$ hard margin, $+$ excursions

Home Page

Title Page

Contents

Page 35 of 41

Go Back

Full Screen

Close

Quit

- $\boxed{L(\mathbf{w}, b, \mathbf{a}, \boldsymbol{\xi}, \boldsymbol{\mu}) \stackrel{\triangle}{=} \frac{1}{2}||\mathbf{w}||^2 -}$ $\boxed{\sum_{i=1}^{N} a_i\{t_i\ [\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) + b] - 1 + \xi_i\} + C\sum_{1=1}^{N} \xi_i - \sum_{i=1}^{N} \mu_i \xi_i}$
- $\{a_i\}$, $\{\xi_i\}$: 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'

Home Page

Title Page

Contents

Page 36 of 41

Go Back

Full Screen

Close

Quit

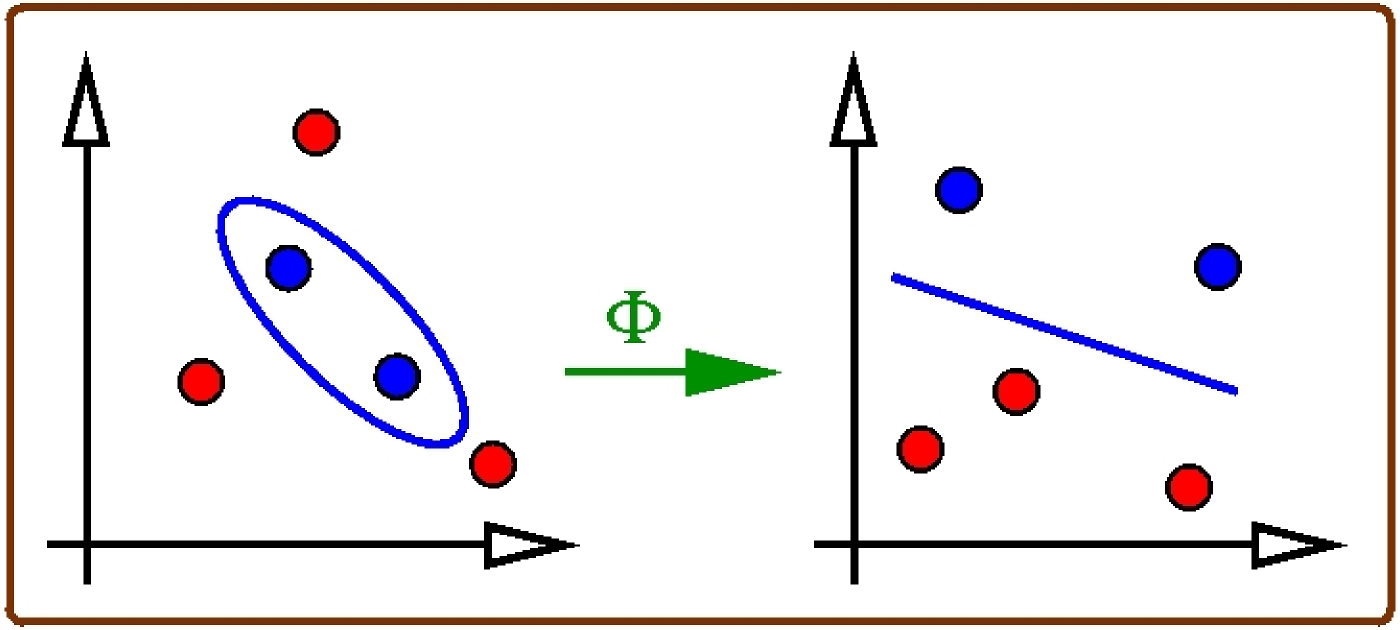
Kernels: Hard Nuts to Crack

http://www.kangarooblue.com/images/fragranceoils/coconut.jpg

- Kernels used in Classification/Regression
- [1964: Aizerman *et al.*], Soviet Control theorists
- [1992: Boser, Guyon, Vapnik]: Hard-M SVM

- Top-down Genesis: (from the SVM Theory):
  - ∗ Many linear models: recast into 'dual' form
  - ∗ Predictors (classification/regression): 'lin combo' of a kernel fn evaluated at (some) training pts
  - ∗ Of interest: models, non-lin feature space mapping, $\mathbf{x} \rightarrow \boldsymbol{\phi}(\mathbf{x})$. e.g., $\tilde{\mathbf{x}}_{3\times 1} = \boldsymbol{\phi}(\mathbf{x}_{2\times 1})$, $\boldsymbol{\phi}(\cdot)$: vector

Kernels: Hard Nuts to Crack



- Basic Idea: Use Φ to map the original patterns in $\mathscr{X}$ into a higher dimensional feature space $\mathscr{H}$
- Draw a separating hyperplane with max margin
- This corresponds to a non-linear decision boundary in the original pattern space
- Kernel trick: using $k(\cdot,\cdot)$, get separating hyperplane without explicitly using Φ to get $\mathscr{H}$

- $k(\mathbf{x}_i, \mathbf{x}_j) \stackrel{\triangle}{=} \boldsymbol{\phi}^T(\mathbf{x}_i)\boldsymbol{\phi}(\mathbf{x}_j) = \boldsymbol{\phi}^T(\mathbf{x}_j)\boldsymbol{\phi}(\mathbf{x}_i)$
- $\mathbf{x}_i$, $\mathbf{x}_j$: vecs in original space, $\boldsymbol{\phi}(\cdot)$: x'fm'd space
- $k(\mathbf{x}_i, \mathbf{x}_j)$: scalar, symmetric, 1st term: fn of $\mathbf{x}_i$, 2nd term: *same* fn of $\mathbf{x}_j$