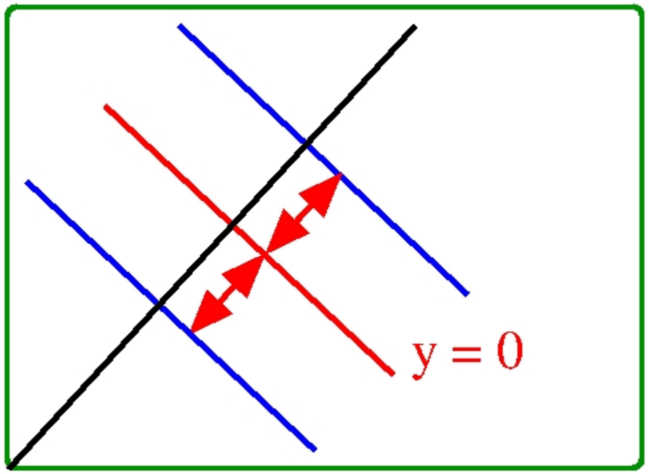
- margin: $\triangleq$ min dist frm a point in either class: sym

Interpretation of: $\arg\max_{\mathbf{w},b}\{\frac{1}{||\mathbf{w}||}\min_i\{t_i\,[\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i)+b]\}\}$

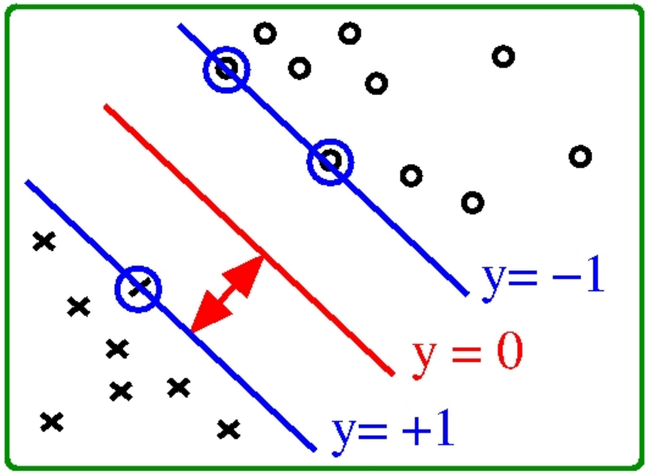
$$= \underbrace{\arg\max_{\mathbf{w},b}\{\underbrace{\min_i \underbrace{\frac{\{t_i \overbrace{[\mathbf{w}^T \overbrace{\boldsymbol{\phi}(\mathbf{x}_i)}^{or,\ \mathbf{x}_i} + b]}^{\text{eqn of line}_{y(\mathbf{x}_i)=0}}\}}{||\mathbf{w}||}}_{\text{dist of } \mathbf{x}_i \text{ from } y(\mathbf{x}_i)=0}}_{\triangleq\ margin}\}}_{\text{to maximise this margin}}$$

- Margin = $\pm\frac{1}{||\mathbf{w}||}$: particularly elegant
- This is just a scaling. Scaling $\mathbf{w}$ and b by κ leave the margin unchanged (shown later)

Three Men in a Boat: Three Lines, Eqns



- $y = 0$: decision boundary
- Implicit form
  $y(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i + b = 0$
- Slope: $\mathbf{w}$, dist from origin $\frac{b}{||\mathbf{w}||}$
- $||$ lines: same slope, diff dist
- 2 lines normalised dist $\frac{1}{||\mathbf{w}||}$
- ‘Near line’: same slope, closer than dec boundary
- closer by $\frac{1}{||\mathbf{w}||}$: from origin: $\frac{b}{||\mathbf{w}||} - \frac{1}{||\mathbf{w}||}$:coeff = $(b-1)$
- $\mathbf{w}^T \mathbf{x}_i + (b-1) = 0 \implies \mathbf{w}^T \mathbf{x}_i + b = +1 \implies y = +1$
- ‘Far line’: same slope, farther than dec boundary
- farther by $\frac{1}{||\mathbf{w}||}$: from origin: $\frac{b}{||\mathbf{w}||} + \frac{1}{||\mathbf{w}||}$:coeff = $(b+1)$
- $\mathbf{w}^T \mathbf{x}_i + (b+1) = 0 \implies \mathbf{w}^T \mathbf{x}_i + b = -1 \implies y = -1$

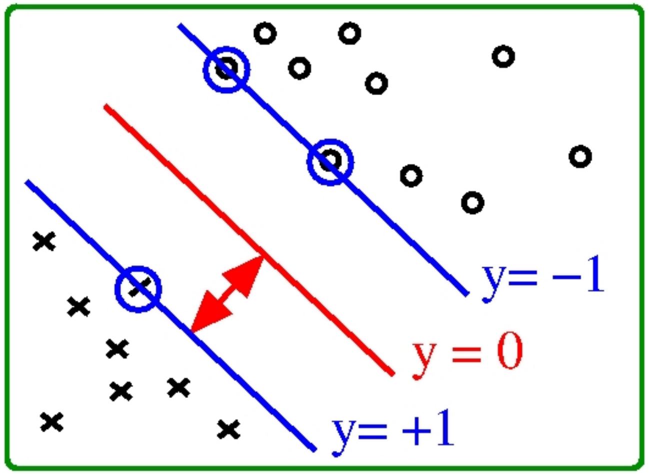


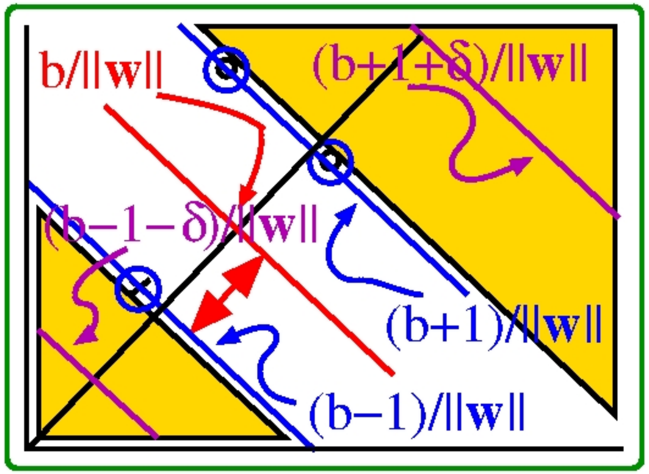
- Maximum margin solution:

$\arg\max_{\mathbf{w},b}\{\frac{1}{||\mathbf{w}||}\min_i\{t_i\ [\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i)+b]\}\}$

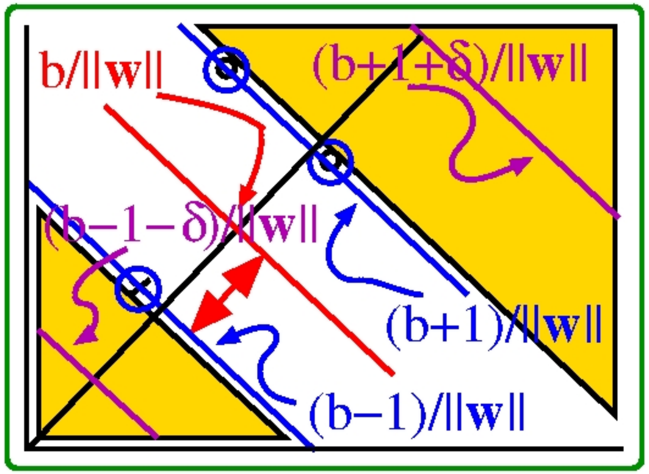
- 'min' comes from 'margin'
- Find $\mathbf{w}, b$ to max the margin
- Now let us look at $y = \pm 1$
- Consider $\boldsymbol{\phi}(\mathbf{x}_i) = \mathbf{x}_i$ (simplicity, no feature xform)
- $y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$ is a hyperplane/line in $\mathbf{x}-$ space
- $\mathbf{w}$ measures the slope/inclination. Why?
- $w_2x_2 + w_1x_1 + b = 0 : \frac{x_2}{(-b/w_2)} + \frac{x_1}{(-b/w_1)} = 1$
- $\frac{b}{||\mathbf{w}||}$: distance from the origin. Vary b: $||$ lines
- If somehow we know the direction $\mathbf{w}$, fit a red line equidistant from the two lines: decision boundary
- How do we know? Oracle/QP solver for $\mathbf{w}$ (& b)
- The distance of a point from the decision boundary is unchanged on a scaling of $\mathbf{w}$ & b by κ each

- The distance of a point from the decision boundary is unchanged on a scaling of $\mathbf{w}$ & b by κ each
- $= \frac{t_i\ \kappa \mathbf{w}^T \mathbf{x}_i + \kappa b}{\kappa ||\mathbf{w}||} = \frac{t_i\ \mathbf{w}^T \mathbf{x}_i + b}{||\mathbf{w}||}$ (property)
- 'Nice' formulation: Consider total margin = $2/||\mathbf{w}||$





- Maximum margin solution: $\arg\max_{\mathbf{w},b}\{\frac{1}{||\mathbf{w}||}\min_i\{t_i\ \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) + b\}\}$
- 'min' comes from 'margin'
- Find $\mathbf{w}, b$ to max the margin
- Now, the $y = \pm 1$ part:
- 2 blue lines @dist$\pm\frac{1}{||\mathbf{w}||} \implies$
- coeff ± 1: $\mathbf{w}^T\mathbf{x} + (b \pm 1) = 0$
- $y = \mathbf{w}^T\mathbf{x} + b = \mp 1$. b +ve
- 'near' line: $y = \mathbf{w}^T\mathbf{x} + b = +1$
- 'far' line: $y = \mathbf{w}^T\mathbf{x} + b = -1$



- We want the golden regions: 2-class data, well-separated
- Consider a magenta line to the right of the blue 'far' line
- Consider 4 dists from origin
- $\frac{b-1}{||\mathbf{w}||}$, $\frac{b}{||\mathbf{w}||}$, $\frac{b+1}{||\mathbf{w}||}$, $\frac{b+1+\delta}{||\mathbf{w}||}$. Last line: $\mathbf{w}^T\mathbf{x}+b=-(1+\delta)$
- Hence for this region $\mathbf{w}^T\mathbf{x}+b<-1$. ||ly the other
- 2 regions: $\mathbf{w}^T\mathbf{x}+b<-1$ & $\mathbf{w}^T\mathbf{x}+b>+1$ $(t_i=\mp 1)$
- $t_i=-1:\mathbf{w}^T\mathbf{x}+b<-1$ & $t_i=+1:\mathbf{w}^T\mathbf{x}+b>+1$
- Generalised Canonical Rep^n: $t_i\,[\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x})+b]>+1$
- Recap: $\boldsymbol{\phi}(\mathbf{x})$ is a feature space xform/kernel fn, for a linear decision boundary in xform space
- SVs: closest to d'boundary vis-a-vis margin
- Optimal margin: linear combo of SVs

Home Page

Title Page

Contents

◀◀ ▶▶

◀ ▶

Page 11 of 41

Go Back

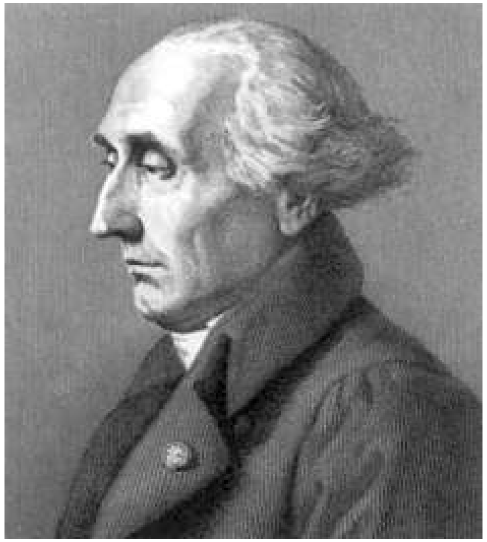
Full Screen

Close

Quit

- Max margin: $\arg\max_{\mathbf{w},b}\{\frac{1}{||\mathbf{w}||}\min_i\{t_i\,[\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i)+b]\}\}$
- $\min_i\{t_i\,[\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i)+b]\}$: margin. So opt: $\max\frac{1}{||\mathbf{w}||}$
- $\max\frac{1}{||\mathbf{w}||}\equiv\min||\mathbf{w}||\equiv\min\frac{1}{2}||\mathbf{w}||^2$. $\frac{1}{2}$: convenience in derivative, square: gets rid of the root in $||\mathbf{w}||$
- $\arg\min_{\mathbf{w},b}\frac{1}{2}||\mathbf{w}||^2$ subject to $t_i\,[\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i)+b]>1,\forall\mathbf{x}_i$
- Quad prog, subject to linear ineq constr.s: $\mathcal{O}(M^3)$
- $L(\mathbf{w},b,\mathbf{a})\stackrel{\triangle}{=}\frac{1}{2}||\mathbf{w}||^2-\sum_{i=1}^{N}a_i\{t_i\,[\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i)+b]-1\}$
- $\frac{1}{2}||\mathbf{w}||^2$: to min, $t_i\,[\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i)+b]-1>0$: sep to max
- $\min(L)$: $||\mathbf{w}||\geq 0$; take max terms neg; constr ≥ 0; lin combo coeffs $a_i\geq 0$: L'mults; $\mathbf{a}=[a_1\ldots a_N]^T$
- Lagrange multipliers: ONE function to max/min, subject to a set of equality/inequality constraints
- $\frac{\partial L}{\partial\mathbf{w}^T}=\mathbf{0}$, $\frac{\partial L}{\partial b}=0$

Intertwined Histories



J.-L. Lagrange
[1736-1813]

A. Lavoisier
[1743-1794]

J.-B. J. Fourier
[1768-1830]

https://upload.wikimedia.org/wikipedia/commons/1/19/Lagrange_portrait.jpg

https://upload.wikimedia.org/wikipedia/commons/4/44/Lavoisier-statue.jpg

https://upload.wikimedia.org/wikipedia/commons/a/aa/Fourier2.jpg

- $\frac{\partial L}{\partial \mathbf{w}^T} = \frac{1}{2} 2\mathbf{w} - \sum_{i=1}^{N} a_i\, t_i\, \boldsymbol{\phi}(\mathbf{x}_i) = 0$: $\mathbf{w} = \sum_{i=1}^{N} a_i\, t_i\, \boldsymbol{\phi}(\mathbf{x}_i)$
- $\frac{\partial L}{\partial b} = 0$: $\sum_{i=1}^{N} a_i\, t_i = 0$
- Under these constraints, what is $L(\mathbf{w}, b, \mathbf{a})$?
- = $\frac{1}{2}\mathbf{w}^T\mathbf{w}$ - $\sum_{i=1}^{N} a_i\, t_i\, \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i)$ - $\sum_{i=1}^{N} a_i\, t_i\, b$ + $\sum_{i=1}^{N} a_i$
- 1st term = $\frac{1}{2}\mathbf{w}^T\mathbf{w} = \frac{1}{2}(\sum_{i=1}^{N} a_i\, t_i\, \boldsymbol{\phi}^T(\mathbf{x}_i))(\sum_{j=1}^{N} a_j\, t_j\, \boldsymbol{\phi}(\mathbf{x}_j))$
- = $\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i\, a_j\, t_i\, t_j\, \boldsymbol{\phi}^T(\mathbf{x}_i)\boldsymbol{\phi}(\mathbf{x}_j)$
- $k(\mathbf{x}_i, \mathbf{x}_j) \triangleq \boldsymbol{\phi}^T(\mathbf{x}_i)\boldsymbol{\phi}(\mathbf{x}_j) = \boldsymbol{\phi}^T(\mathbf{x}_j)\boldsymbol{\phi}(\mathbf{x}_i)$
- 1st term = $\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i\, a_j\, t_i\, t_j\, k(\mathbf{x}_i, \mathbf{x}_j)$
- 2nd term = $\sum_{i=1}^{N} a_i\, t_i\, (\sum_{j=1}^{N} a_j\, t_j\, \boldsymbol{\phi}^T(\mathbf{x}_j))\boldsymbol{\phi}(\mathbf{x}_i))$
- = $\sum_{i=1}^{N}\sum_{j=1}^{N} a_i\, a_j\, t_i\, t_j\, k(\mathbf{x}_i, \mathbf{x}_j)$
- 3rd term = $b\sum_{i=1}^{N} a_i\, t_i \;= 0$

Home Page
Title Page
Contents
Page 13 of 41
Go Back
Full Screen
Close
Quit

Home Page

Title Page

Contents

Page 14 of 41

Go Back

Full Screen

Close

Quit

$$L(\mathbf{w}, b, \mathbf{a}) = \widetilde{L}(\mathbf{a}) = \sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i\, a_j\, t_i\, t_j\, k(\mathbf{x}_i, \mathbf{x}_j)$$

- subject to $a_i \geq 0$ (Lagrange multipliers) & $\sum_{i=1}^{N} a_i\, t_i = 0$: Dual Formulation: no $\mathbf{w}$,b: at opt
- constrained $L(\mathbf{w}, b, \mathbf{a}) \rightarrow$ constrained $\widetilde{L}(\mathbf{a})$ Dual
- Weird? Original $\arg\min_{\mathbf{w},b} \frac{1}{2}\mathbf{w}^T\mathbf{w}$: M-dim, $\mathscr{O}(M^3)$
- Dual $\arg\min_{\mathbf{a}} \sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i\, a_j\, t_i\, t_j\, k(\mathbf{x}_i, \mathbf{x}_j)$: N-dim problem
- $D = (M-1)-$dim formulation: $y(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}) + b$
- $D = 2$: $y(x_2, x_1) = [w_2\ w_1] \begin{bmatrix} \phi_2(x_2, x_1) \\ \phi_1(x_2, x_1) \end{bmatrix} + b$
- $y(\mathbf{x}) = [w_2\ w_1] \begin{bmatrix} \phi_2(\mathbf{x}) \\ \phi_1(\mathbf{x}) \end{bmatrix} + b = w_2 x_2 + w_1 x_1 + b$ (omit $\boldsymbol{\phi}$)
- 2-D line: coeffs 0 ($w_0 = b$) to $M-1$, 2-D weights $\mathbf{w}$ & one b: $M = D+1$ params typically $< N$ (# points)
- Kernel: transform data to a higher dim space