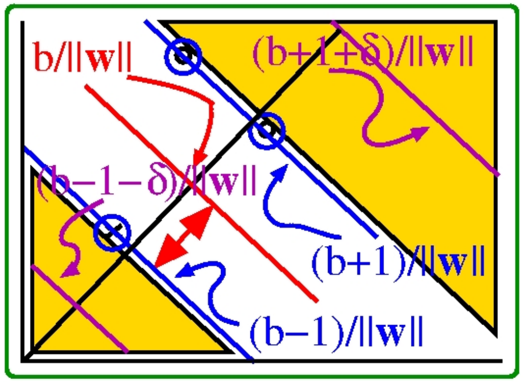
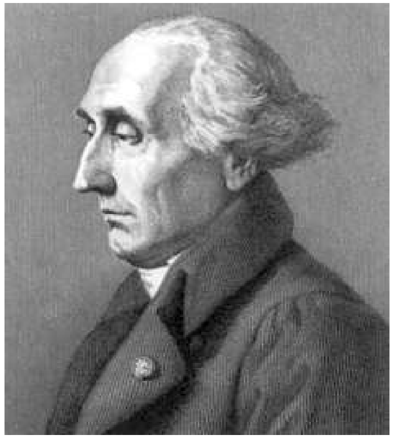



- We want the golden regions: 2-class data, well-separated
- Consider a magenta line to the right of the blue 'far' line
- Consider 4 dists from origin
- $\frac{b-1}{||\mathbf{w}||}$, $\frac{b}{||\mathbf{w}||}$, $\frac{b+1}{||\mathbf{w}||}$, $\frac{b+1+\delta}{||\mathbf{w}||}$. Last line: $\mathbf{w}^T\mathbf{x}+b=-(1+\delta)$
- 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

- 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: $\mathcal{O}(M^3)$
- $L(\mathbf{w},b,\mathbf{a})\triangleq\frac{1}{2}||\mathbf{w}||^2-\sum_{i=1}^N a_i\{t_i\,[\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i)+b]-1\}$
- $\frac{1}{2}||\mathbf{w}||^2$: 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 $\geq 0$; 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

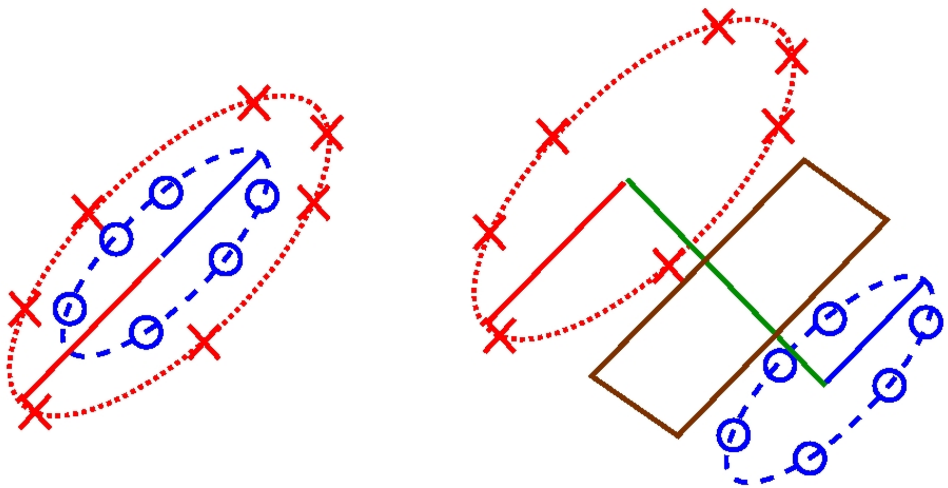
  

- $\frac{\partial L}{\partial \mathbf{w}^T} = \frac{1}{2} 2\mathbf{w} - \sum_{i=1}^{N} a_i \, t_i \, \boldsymbol{\phi}(\mathbf{x}_i) = 0$: $\mathbf{w} = \sum_{i=1}^{N} a_i \, t_i \, \boldsymbol{\phi}(\mathbf{x}_i)$
- $\frac{\partial L}{\partial b} = 0$: $\sum_{i=1}^{N} a_i \, t_i = 0$
- Under these constraints, 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 = $\frac{1}{2}\mathbf{w}^T\mathbf{w} = \frac{1}{2}(\sum_{i=1}^{N} a_i \, t_i \, \boldsymbol{\phi}^T(\mathbf{x}_i))(\sum_{j=1}^{N} a_j \, t_j \, \boldsymbol{\phi}(\mathbf{x}_j))$
- = $\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i \, a_j \, t_i \, t_j \, \boldsymbol{\phi}^T(\mathbf{x}_i)\boldsymbol{\phi}(\mathbf{x}_j)$
- $k(\mathbf{x}_i, \mathbf{x}_j) \stackrel{\triangle}{=} \boldsymbol{\phi}^T(\mathbf{x}_i)\boldsymbol{\phi}(\mathbf{x}_j) = \boldsymbol{\phi}^T(\mathbf{x}_j)\boldsymbol{\phi}(\mathbf{x}_i)$
- 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 = $b \sum_{i=1}^{N} a_i \, t_i \; = 0$

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

- 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, $\mathcal{O}(M^3)$
- 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



- The 2 classes (left) have the same centre in 2-D
- Separable by a circle, not a lin decision boundary
- Transform it to 3-D, the third coord = radius
- $[x, y, r]$: larger circle floats up, separating plane
- Kernel function: to higher dim, hope: lin boundary
- Kernel trick: May not need to transform
- Comps in inner product space $\boldsymbol{\phi}^T(\mathbf{x})\boldsymbol{\phi}(\mathbf{x}) = \mathbf{x}^T\mathbf{x}$
- Philosophically: like energy in Parseval's theorem