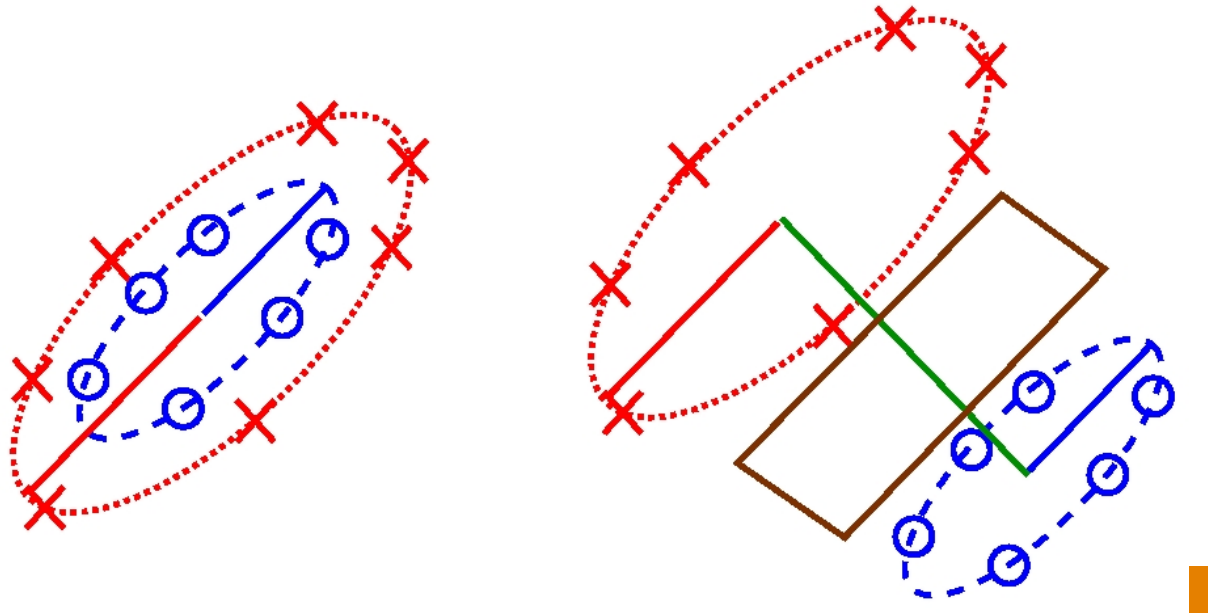
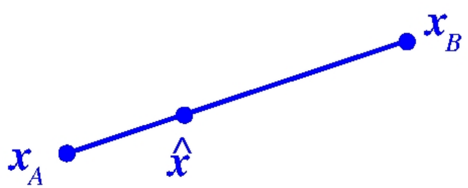


- The 2 classes (left) have the same centre in 2-D
- Separable by a circle, not a lin decision boundary
- Transform it to 3-D, the third coord = radius
- $[x, y, r]$: larger circle floats up, separating plane
- Kernel function: to higher dim, hope: lin boundary
- Kernel trick: May not need to transform
- 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

Lagrange Multipliers & KKT

- Equality, Inequality, Both Equality & Inequality
- opt $f(\mathbf{x})$ subject to $g(\mathbf{x}) = 0$. opt is $\max$ or $\min$
- $\mathbf{x}$ is $D-$dim, $g(\mathbf{x}) = 0$: surface in $(D-1)-$dim
- e.g., 2-D space, constraint: line (1-D) $\mathbf{w}^T\mathbf{x} + b = 0$
- Find opt $f(\mathbf{x})$ on surface (lin/non-lin) $g(\mathbf{x}) = 0$
- Not imp: $g(\mathbf{x} + \boldsymbol{\varepsilon}) \approx g(\mathbf{x}) + \boldsymbol{\varepsilon}^T \nabla g(\mathbf{x})$. $\boldsymbol{\varepsilon}^T \nabla g(\mathbf{x}) \approx 0$
- $\boldsymbol{\varepsilon}$ on the surface $\perp$ $\nabla g(\mathbf{x})$ normal to it
- Lagrange: $L(\mathbf{x}, \lambda) \stackrel{\triangle}{=} f(\mathbf{x}) + \lambda g(\mathbf{x})$, lin combo $\lambda \geq 0$
- 

  - Slider b/w 2 scalars/vectors
  - value = either endpoint, or in b/w
- Spl case $\lambda \in [0,1]$: probability connotation!
- $\lambda < 0$: ext div of line by a pt: $\lambda \geq 0$ not restrictive!

- params $\mathbf{x}, \lambda$. $\frac{\partial L}{\partial param} = 0$. $\frac{\partial L}{\partial \lambda} = 0$: constraint $g(\mathbf{x}) = 0$
- L's trick: 'undetermined multipliers', why eval λ
- $\nabla_{\mathbf{x}} f(\mathbf{x}) + \lambda \nabla_{\mathbf{x}} g(\mathbf{x}) = 0 : \nabla_{\mathbf{x}} f(\mathbf{x}) = -\lambda \nabla_{\mathbf{x}} g(\mathbf{x})$
- If $g(\mathbf{x}) = 0$, isn't $L(\mathbf{x}, \lambda) = f(\mathbf{x})$? Only @the opt!

Extension to one Inequality Constraint

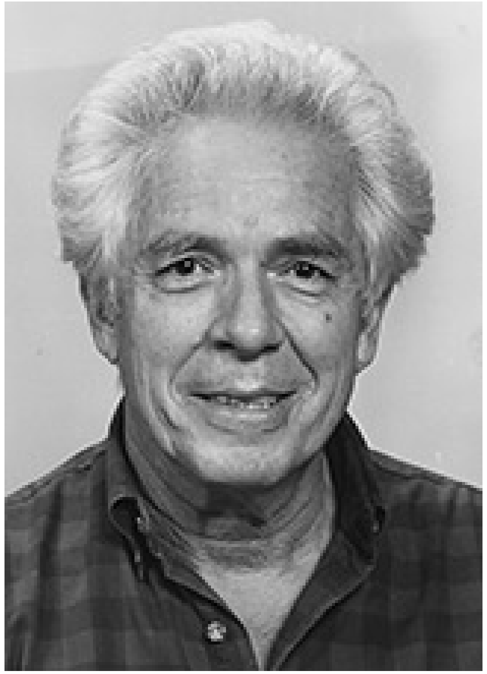
- $\max f(\mathbf{x})$ subject to $g(\mathbf{x}) \geq 0 \implies$
- $\max f(\mathbf{x})$ on the surface $g(\mathbf{x}) = 0$ & on one side
- 2 cases: on surface $\{g(\mathbf{x}) = 0\}$& 1 side $\{g(\mathbf{x}) > 0\}$
- Case 1: off the surface $g(\mathbf{x})$ has no role
- $\lambda = 0$: $L(\mathbf{x}, \lambda) \stackrel{\triangle}{=} f(\mathbf{x}) + \lambda g(\mathbf{x}) = f(\mathbf{x})$, $\nabla_{\mathbf{x}} L = \nabla_{\mathbf{x}} f(\mathbf{x})$
- Case 2: on surface, L = linear combo of $f(\mathbf{x}), g(\mathbf{x})$
- On the surface: not at an end-point: $\lambda > 0$
- $\nabla_{\mathbf{x}} L = \nabla_{\mathbf{x}} f(\mathbf{x}) + \lambda \nabla_{\mathbf{x}} g(\mathbf{x}) = 0$: $\nabla_{\mathbf{x}} f(\mathbf{x}) = -\lambda \nabla_{\mathbf{x}} g(\mathbf{x})$

The Karush-Kuhn-Tucker Conditions

1. $\lambda \geq 0$: Linear combo 2. $g(\mathbf{x}) \geq 0$: Constraint

3. $\lambda g(\mathbf{x}) = 0$: off-surface $\lambda = 0$, on-surface $g(\mathbf{x}) = 0$

W. Karush
(1939 U Chicago)
Masters Thesis
[1917-1997]



H. W. Kuhn
1951 paper
[1925-2014]

A. W. Tucker
1951 paper
[1905-1995]

Tucker's famous students include J. F. Nash, M. Minsky

https://s3-us-west-2.amazonaws.com/find-a-grave-prod/photos/2012/236/95844065_134584501059.jpg
https://upload.wikimedia.org/wikipedia/en/b/b4/Harold_W._Kuhn.jpg
https://upload.wikimedia.org/wikipedia/en/4/4a/Albert_W._Tucker.gif

Home Page

Title Page

Contents

Page 19 of 41

Go Back

Full Screen

Close

Quit

Max/Min with one Inequality Constraint

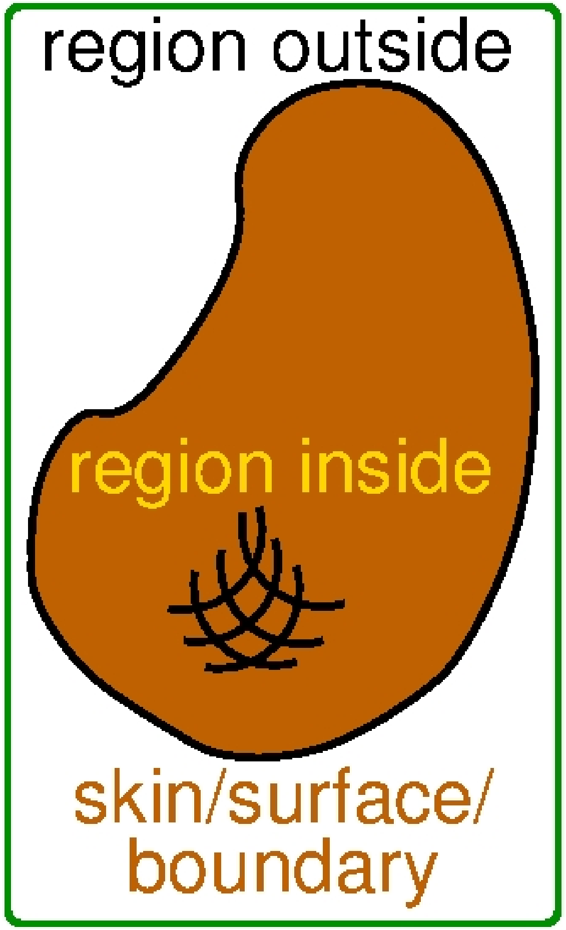
- To maximise $f(\mathbf{x})$ subject to $g(\mathbf{x}) \geq 0$
- $\implies$ Maximise $L(\mathbf{x}, \lambda) \stackrel{\triangle}{=} f(\mathbf{x}) + \lambda g(\mathbf{x}), \lambda \geq 0$
- To minimise $f(\mathbf{x})$ subject to $g(\mathbf{x}) \geq 0$
- $\implies$ Minimise $L(\mathbf{x}, \lambda) \stackrel{\triangle}{=} f(\mathbf{x}) - \lambda g(\mathbf{x}), \lambda \geq 0$
- $g(\mathbf{x}) \leq 0$ tackled as $\widetilde{g}(\mathbf{x}) \geq 0, \widetilde{g}(\mathbf{x}) = -g(\mathbf{x})$

Multiple Equality/Inequality Constraints

- Max $f(\mathbf{x})$ subject to $g_j(\mathbf{x}) = 0, h_k(\mathbf{x}) \geq 0$: $\lambda_j, \mu_k \geq 0$

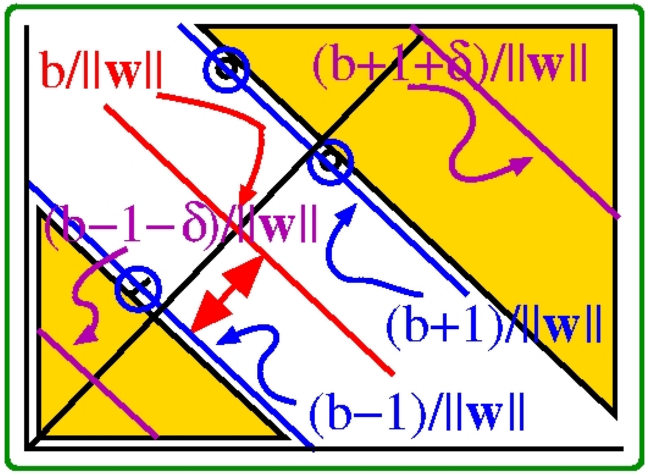
$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \stackrel{\triangle}{=} f(\mathbf{x}) + \sum_{j=1}^{J} \lambda_j g_j(\mathbf{x}) + \sum_{k=1}^{K} \mu_k h_k(\mathbf{x})$$

The KKT Conditions and SVMs



- The 'surface' $g(\mathbf{x}) = 0$ divides the space into two parts, one 'inside' and one 'outside'. In any one $g(\mathbf{x}) > 0$, and in the other, $g(\mathbf{x}) < 0$.
- $g(\mathbf{x}) \geq 0$ indicates the surface and a region (outside the potato, say).
- $\min \ f(\mathbf{x})$ subject to $g(\mathbf{x}) \geq 0$

  $\equiv \min L(\mathbf{x}, \lambda) \stackrel{\triangle}{=} f(\mathbf{x}) - \lambda g(\mathbf{x})$
- Physical Significance: min $f(\mathbf{x})$ in the region $g(\mathbf{x}) \geq 0$. 2 places: skin $g(\mathbf{x}) = 0$; region $g(\mathbf{x}) > 0$
- To derive the KKT conditions for a hard-margin SVM, from Lagrange Multipliers
- Graphical derivation: the two potato cases. Just as (mathematically) nutritious



- **SVM constraints** $t_i y_i \geq 1$
- Recap: golden regions from representative points on lines || decision boundary
- db: $\mathbf{w}^T\mathbf{x}_i + b = 0$: $y(\mathbf{x}_i) = y_i = 0$
- The 'near' golden region: closer than the 'near' blue line (normalised dist 1 closer than the db)
- 'Near' golden region, magenta line: $\mathbf{w}^T\mathbf{x}_i + (b - (1+\delta)) = 0$: $\mathbf{w}^T\mathbf{x}_i + b = 1 + \delta$: $y_i \geq +1$ (as $\delta \geq 0$)
- Here, $t_i = +1$ (convenience): $t_i y_i \geq 1$
- 'Far' golden region, magenta line: $\mathbf{w}^T\mathbf{x}_i + (b + (1+\delta)) = 0$: $\mathbf{w}^T\mathbf{x}_i + b = -(1+\delta)$: $y_i \leq -1$ (as $\delta \geq 0$)
- Here, $t_i = -1$ (convenience): $t_i y_i \geq 1$
- Overall Constraint: $t_i y_i \geq 1$ or, $t_i y_i - 1 \geq 0$

Lagrange	KKT
$\min f(\mathbf{x})$ subj to $g(\mathbf{x}) \geq 0$	$\min \frac{1}{2}\|\|\mathbf{w}\|\|^2$ subj to $t_i y_i \geq 1$
$\min L(\mathbf{x}, \lambda) \triangleq f(\mathbf{x}) - \lambda g(\mathbf{x})$	$\min\ L(\mathbf{w}, b, \mathbf{a}) \triangleq \frac{1}{2}\|\|\mathbf{w}\|\|^2 - \sum a_i[t_i y_i - 1]$
1. $\lambda \geq 0$: Lagrange mults	1. $a_i \geq 0$: Lagrange mults
2. $g(\mathbf{x}) \geq 0$: Constraint 2 cases: $g(\mathbf{x}) > 0$ (region) & $g(\mathbf{x}) = 0$ (skin)	2. $t_i y_i \geq 1$: Constraint 2 cases: $t_i y_i > 1$ (golden) & $t_i y_i = 1$ (margin lines)
3. [imp] $\lambda g(\mathbf{x}) = 0$	3. [imp] $a_i[t_i y_i - 1] = 0$
a. $g(\mathbf{x}) > 0$: region: no effect of constraint: $\lambda = 0$	a. $t_i y_i - 1 > 0$: golden: no effect of constraint: $a_i = 0$
b. $g(\mathbf{x}) = 0$: skin: $\lambda > 0$	b. $t_i y_i = 1$: margin lines, $a_i > 0$: Support Vectors