CSAIL

SVMs

Lecture 1

- Quick SVM Recap
  - Separating hyperplanes: $\vec{w} \cdot \vec{x} + b = 0$
  - margins: distance from point to line
  - Kernels: compute dot product in transformed space, w/o trans. data

-noise

- Quick Linear Algebra Review
  - inner product (dot product)
  - hyperplanes
  - distance of point to line
  - scaling
- Quick Lagrangian Optimization Review
  - equivalence of two methods
  (Do you encode exact constraint and take partials w.r.t. Lagrange multipliers as well?)
  ↳ we will

later

- SVMs
  - Start w/ tutorial
    - perceptron
    - duality
    - Kernels
  - move to notes
    - functional margin
    - geometric margin
    - optimal margin classifier

large margin classifier.

Lecture 2

1.

- Perceptron Recap - intro to duality
- Functional & Geom. margin recap
  - SVM optimization setup

- Lagrange Opt Review - duality
- SVM primal to dual
  - solving for w^* & b^*
- Kernel trick
- non-separability & outliers

The Stata Center, Building 32
32 Vassar Street, Cambridge, Massachusetts 02139-4307
WEB csail.mit.edu

Lagrangian Optimization

Two equivalent methods:

E.g. max $x+y$ s.t. $x^2+y^2=1$

1

- $J(x,y) = (x+y) + \lambda(x^2+y^2)$
- take partials w.r.t. x & y; set to zero to get form

$$\frac{\partial J}{\partial x} = 1 + 2\lambda x = 0 \Rightarrow x = -1/2\lambda$$

$$\frac{\partial J}{\partial y} = 1 + 2\lambda y = 0 \Rightarrow y = -1/2\lambda$$

- Plug into constraint to find value of Lagrange multiplier

$$(-1/2\lambda)^2 + (-1/2\lambda)^2 = 1$$

$$\Rightarrow \lambda = \pm 1/\sqrt{2}$$

- Plug back into form

$$(x,y) = \left(\frac{\sqrt{2}}{2}, \sqrt{2}/2\right)$$

$$\text{or } \left(-\sqrt{2}/2, -\sqrt{2}/2\right)$$

2

- Functional encodes constraints and is a function of Lagrange multipliers

$$J(x,y,\lambda) = (x+y) + \lambda(x^2+y^2-1)$$

- Take partials w.r.t. all functional variables, set to zero

$$\frac{\partial J}{\partial x} = 1 + 2\lambda x = 0 \qquad (x = -1/2\lambda)$$

$$\frac{\partial J}{\partial y} = 1 + 2\lambda y = 0 \qquad (y = -1/2\lambda)$$

$$\frac{\partial J}{\partial \lambda} = x^2+y^2-1 = 0 \quad \Leftarrow \text{ just constraint}$$

- Solve simultaneously

Duality: Solve for primary variables, plug in and obtain function of Lagrange variables

$$J(\lambda) = (-1/2\lambda - 1/2\lambda) + \lambda\left((-1/2\lambda)^2 + (-1/2\lambda)^2 - 1\right)$$

$$= -1/\lambda + \lambda\left(1/2\lambda^2 - 1\right)$$

$$= -1/\lambda + 1/2\lambda - \lambda$$

$$= \frac{-1}{2\lambda} - \lambda$$

Now we have an optimization problem in Lagrange variables —
- take deriv. and set to zero, or
- solve numerically

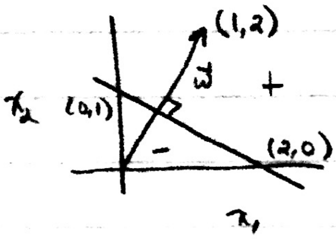
Some Basic Linear Algebra

- Inner product (dot product)

$$\vec{w}\cdot\vec{x} = \langle \vec{w}, \vec{x} \rangle = w^T x = \sum_i w_i \cdot x_i$$

- Vector length $\|\vec{w}\| = \sqrt{\vec{w}\cdot\vec{w}} = \sqrt{\sum_i w_i^2}$
- Hyperplane

$$\vec{w}\cdot\vec{x} + b = 0$$



$x_1 + 2x_2 = 2$

$\vec{w} = (1,2)$

$b = -2$

$\Rightarrow \vec{w}\cdot\vec{x} + b = 0$

- Pos/Neg dist to line

$$\frac{\vec{w}\cdot\vec{x} + b}{\|\vec{w}\|} \text{ – can be pos or neg}$$

if $\vec{w}$ is unit vector, then $\|\vec{w}\| = 1$, so just

$$\vec{w}\cdot\vec{x} + b$$

Note. $10\times$ $x_1 + 2x_2 = 2$ is same line

$\Rightarrow$ $10x_1 + 20x_2 = 20$

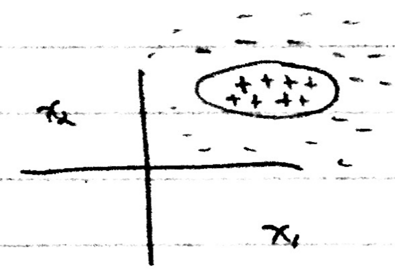
$$\left.\begin{aligned} \vec{w} &= (10, 20) \\ b &= -20 \end{aligned}\right\} \text{scaling}$$

Canonical: often normalize to unit $\perp$ vector

$$\frac{\vec{w}}{\|\vec{w}\|}\cdot\vec{x} + \frac{b}{\|\vec{w}\|} = 0$$

Kernel example

Consider ellipse:



centred at (x_1^c, x_2^c)

$$\frac{(x_1 - x_1^c)^2}{a^2} + \frac{(x_2 - x_2^c)^2}{b^2} \leq 1$$

$$\Rightarrow \frac{1}{a^2} \cdot x_1^2 + \frac{1}{b^2} x_2^2 - \frac{2x_1^c}{a^2} x_1 - \frac{2x_2^c}{b^2} x_2 + \left(\frac{x_1^{c^2}}{a^2} + \frac{x_2^{c^2}}{b^2} - 1\right) \leq 0$$

$$\Rightarrow \vec{w} \cdot \vec{z} + b \leq 0$$ where

$\vec{w} = \left(\frac{1}{a^2}, \frac{1}{b^2}, -\frac{2x_1^c}{a^2}, -\frac{2x_2^c}{b^2}\right)$ ← constant coeff

$\vec{z} = (x_1^2, x_2^2, x_1, x_2)$ ← polynomial kernel

$b = \frac{x_1^{c^2}}{a^2} + \frac{x_2^{c^2}}{b^2} - 1$

↖ constant coeff.