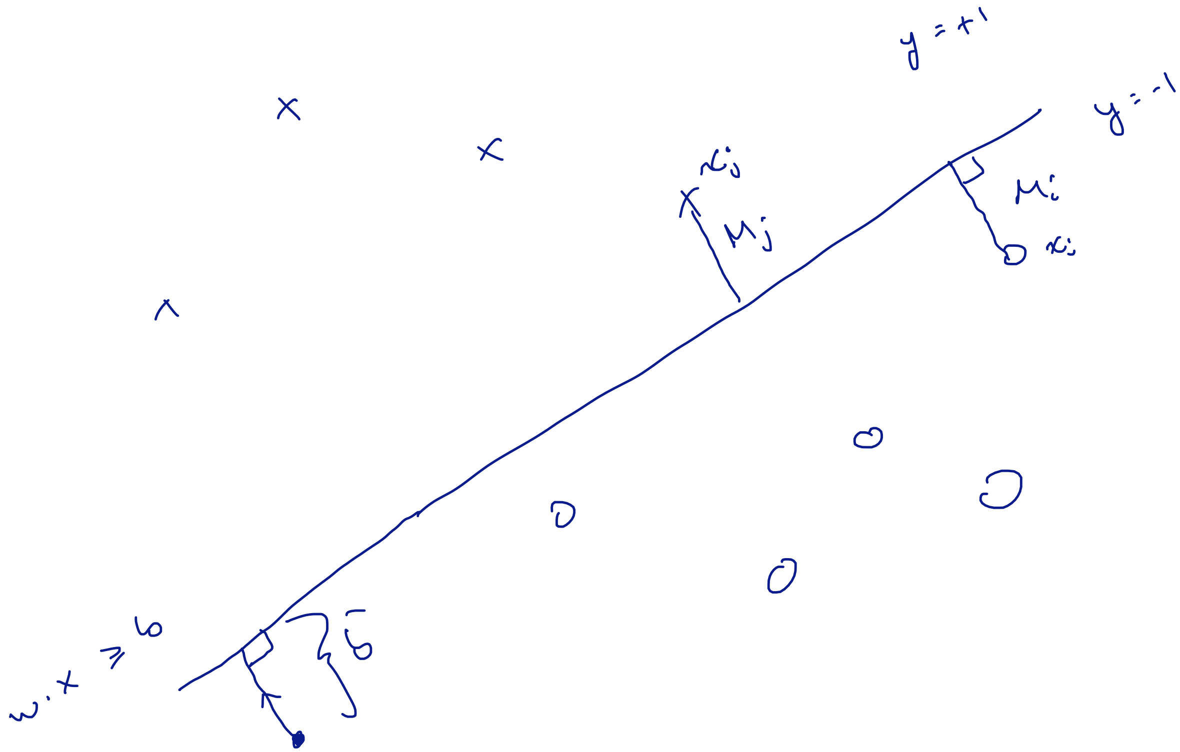
Administrivia

- HW4 out
  - based on feedback survey,
    - fewer questions: 4, but only do 3
    - range of problem types: focus on those that help your understanding
    - split out "spoilers" for Q2
- Midterm
  - mean 65 (out of 95), std dev 11.3
  - ~~back at end of class~~

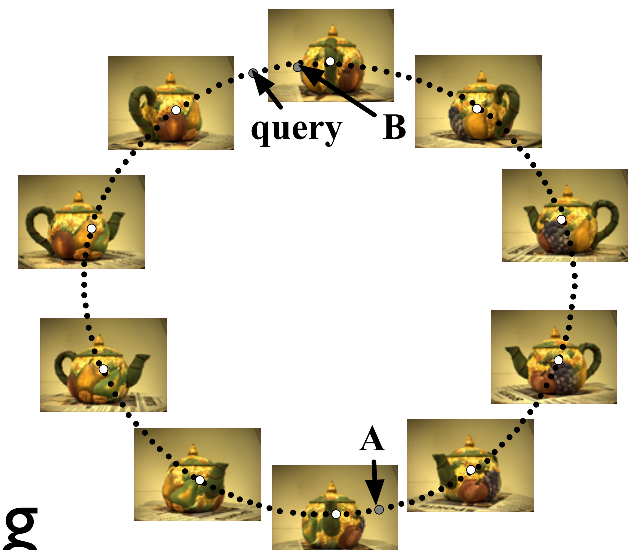
Review

- Cone & QP duality
  - $\min\ c^T x + x^T H x/2 \quad \text{s.t.} \quad Ax + b \in K \quad x \in L$
  - $\max\ -z^T H z/2 - b^T y \quad \text{s.t.} \quad Hz + c - A^T y \in L^* \quad y \in K^*$
- KKT conditions
  - primal: $Ax+b \in K \quad x \in L$
  - dual: $Hz + c - A^T y \in L^* \quad y \in K^*$
  - quadratic: $Hx = Hz$
  - comp. slack: $y^T(Ax+b) = 0 \quad x^T(Hz+c-A^T y) = 0$

Review



Support vector machines



Maximum-variance unfolding

Support vector machines

10-725 Optimization
Geoff Gordon
Ryan Tibshirani

SVM duality

$A = \begin{pmatrix} y_i x_i^T \end{pmatrix}$

- min $\|v\|^2/2 + \Sigma s_i$ s.t. $y_i (x_i^T v - d) \geq 1 - s_i$ $s_i \geq 0$
- min $\underline{v^T v/2 + 1^T s}$ s.t. $Av - yd + s - 1 \geq 0$ $s \geq 0$

$\alpha \geq 0$ $\quad t \geq 0$

$\alpha^T (Av - yd + s - 1) \geq 0$ $\qquad t^T s \geq 0$

$K = AA^T$

$\text{obj} \geq v^T v/2 + 1^T s - \alpha^T (Av - yd + s - 1) - t^T s$

$\geq \min_{v,d,s} [\;\uparrow\;] = -\alpha^T A A^T \alpha/2 + 1^T \alpha$

$\nabla_v : \; 0 = v - A^T \alpha \;\rightarrow\; v = A^T \alpha$

$\nabla_d : \; 0 = \alpha^T y$

$\nabla_s : \; 0 = 1 - \alpha - t \qquad 1 = \alpha + t$

$$\boxed{\begin{array}{l} \max\; 1^T \alpha - \alpha^T K \alpha/2 \;\; \text{st} \\ \alpha^T y = 0 \\ \alpha \geq 0 \\ \alpha \leq 1 \end{array}}$$

Interpreting the dual

- $\max\ 1^T\alpha - \alpha^T K \alpha/2 \quad \text{s.t.}\ \ y^T\alpha = 0 \quad 0 \le \alpha \le 1$

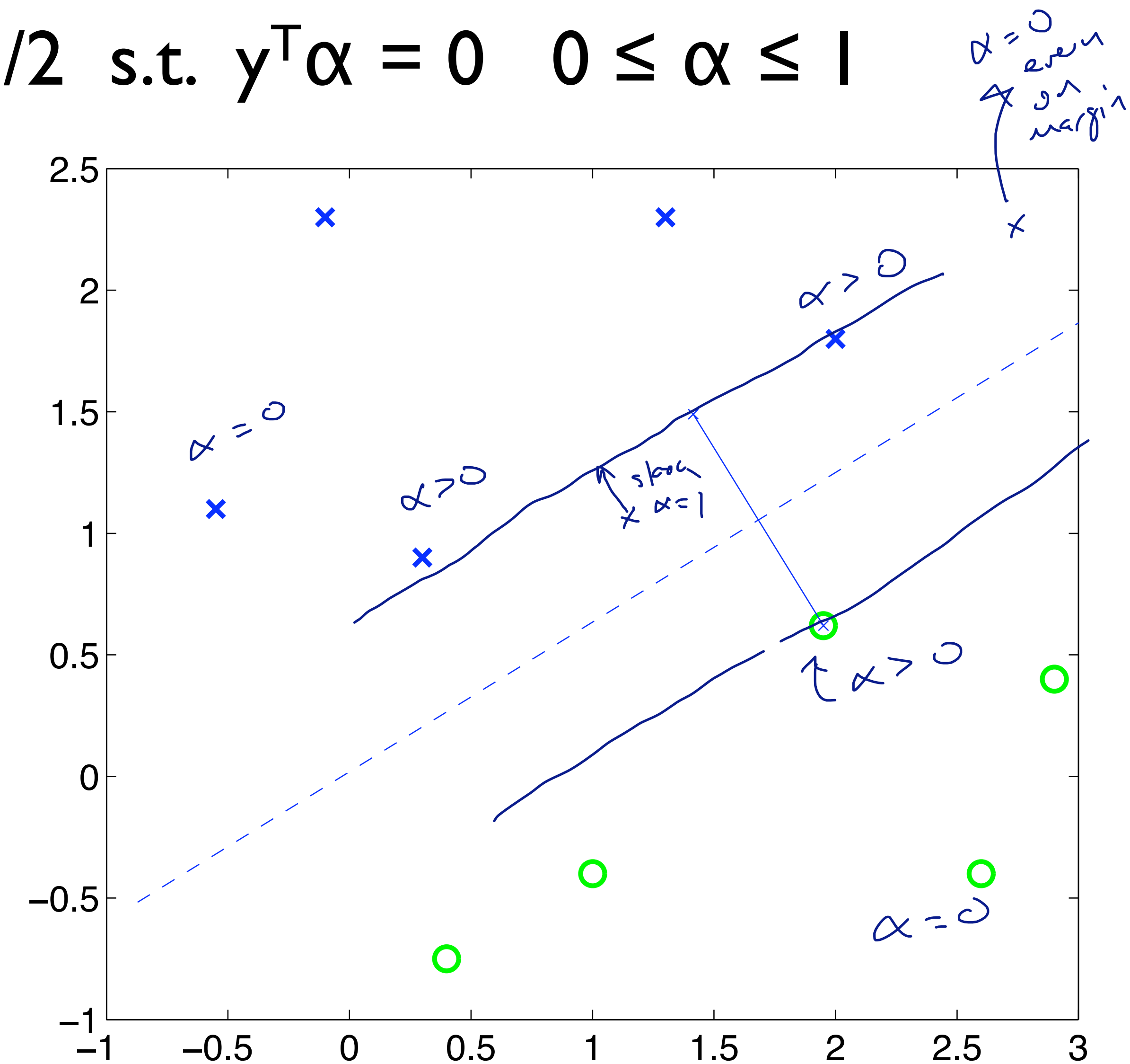
α: example weights

$\alpha>0$: support vectors

$\alpha<1$: slack = 0

$y^T\alpha=0$: $\sum_i y_i \alpha_i = 0$

$$\sum_{+ve} \alpha_i = \sum_{-ve} \alpha_i = \alpha_{tot}$$



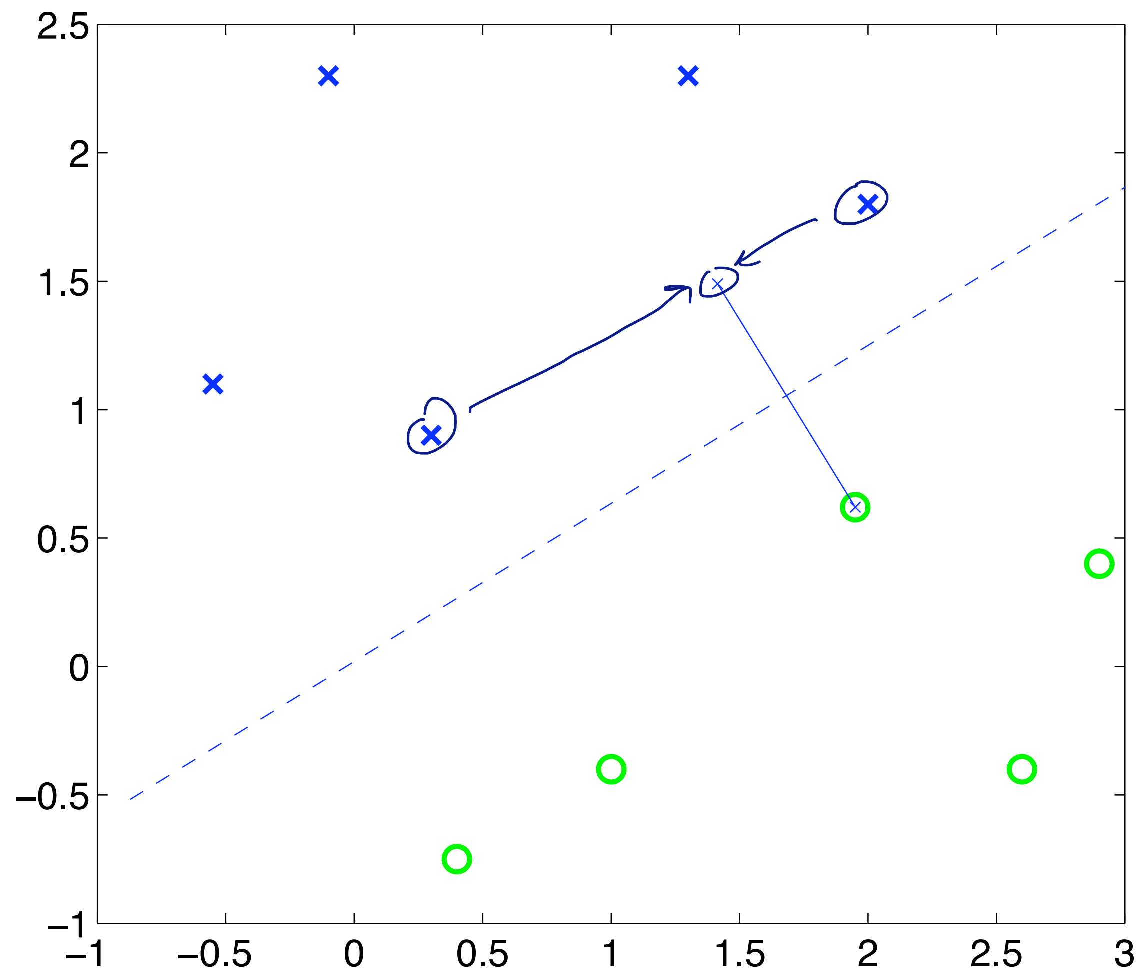
From dual to primal

- max $1^T\alpha - \alpha^T K\alpha/2$ s.t. $y^T\alpha = 0 \quad 0 \le \alpha \le 1$

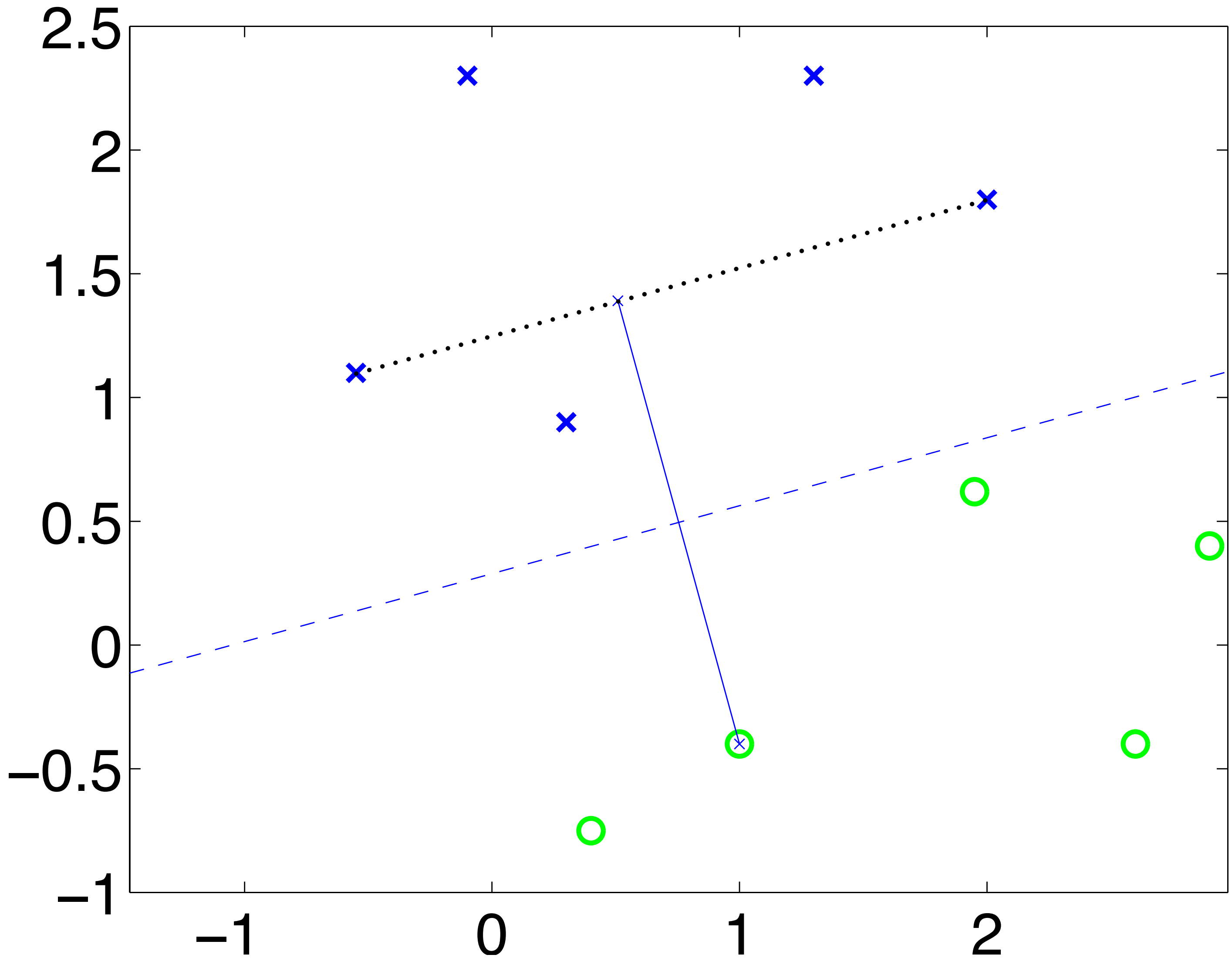
$$\frac{v}{\alpha_{tot}} = A^T\alpha$$
$$= \sum_i y_i x_i \alpha_i$$
$$= \sum_{+ve} \frac{x_i \alpha_i}{\alpha_{tot}} - \sum_{-ve} \frac{x_i \alpha_i}{\alpha_{tot}}$$

$\alpha_i > 0$

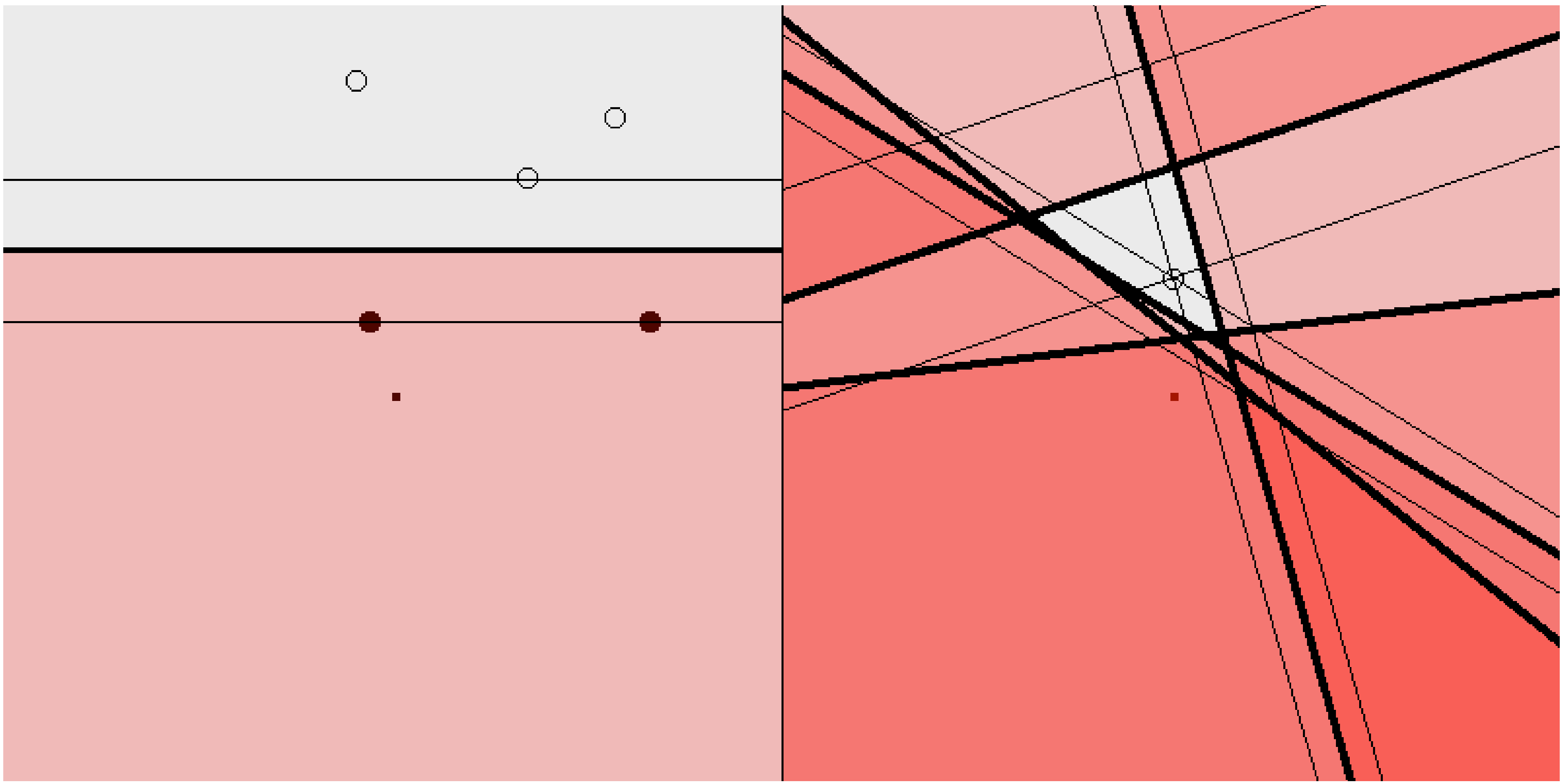
$y_i x_i \cdot v - y_i d = 1 - s_i$



A suboptimal support set



SVM duality: the applet



Why is the dual useful?

$$\max\ 1^T\alpha - \alpha^T K\alpha/2 \ \text{ s.t. }\ y^T\alpha = 0 \quad 0 \le \alpha \le 1$$

- SVM: n examples, m features: $x_i = \phi(u_i) \in R^m$
  - primal: m+1 vars n constrs
  - dual: n vars 1 equality 2n box

The kernel trick

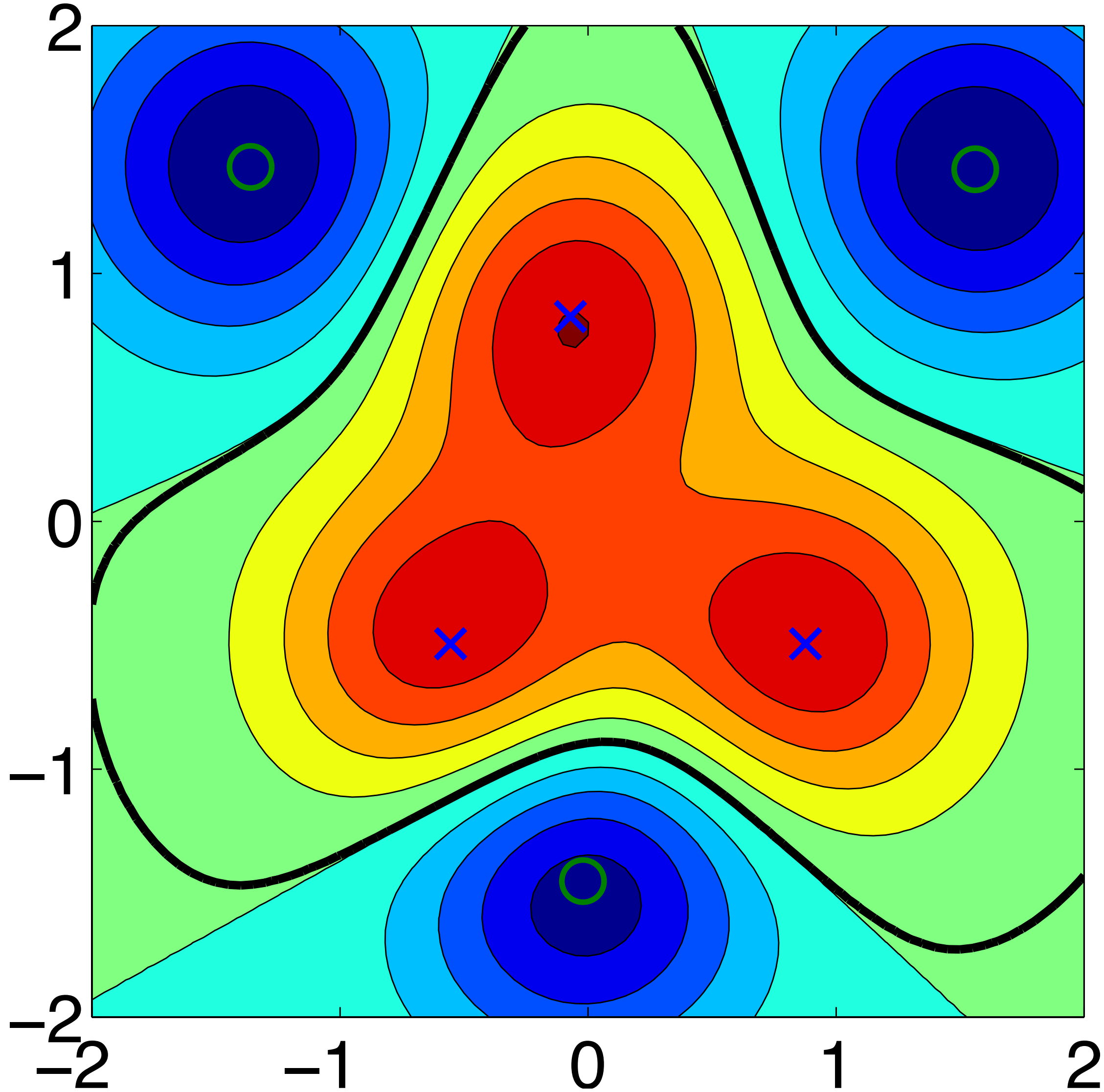
- Don't even need to know features $x_i = \phi(u_i)$, as long as we can compute dot products $x_i^T x_j$
- Matrix of dot products:
  - $K_{ij} =$ $x_i^T x_j = \phi(u_i)^T \phi(u_j) = k(u_i, u_j)$
  - only need subroutine for k (don't care about ϕ)
  - how do we know k works?
    - $K \succeq 0 \quad K = K^T \quad \forall \{u_i\}_{i=1..n}$
    - this is a "positive definite function," aka "Mercer kernel"—∃ many examples

Examples of kernels

- $K(u_i, u_j) = (1 + u_i^T u_j)^d$
  - can represent any degree-d polynomial
  - i.e., decision surface is $p(u) = b$ for degree-d poly p
- $K(u_i, u_j) = (u_i^T u_j)^d$
  - polynomial where all terms have degree exactly d
  - d=1 reduces to original (linear) SVM
- $K(u_i, u_j) = \exp(-||u_i - u_j||^2 / 2\sigma^2)$
  - Gaussian radial basis functions of width σ

Gaussian kernel

$\sigma = 0.5$



Interior-point methods

10-725 Optimization
Geoff Gordon
Ryan Tibshirani

Ball center

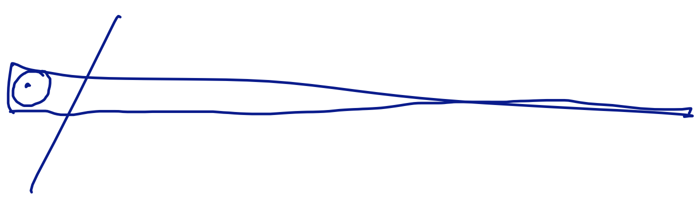
aka Chebyshev center

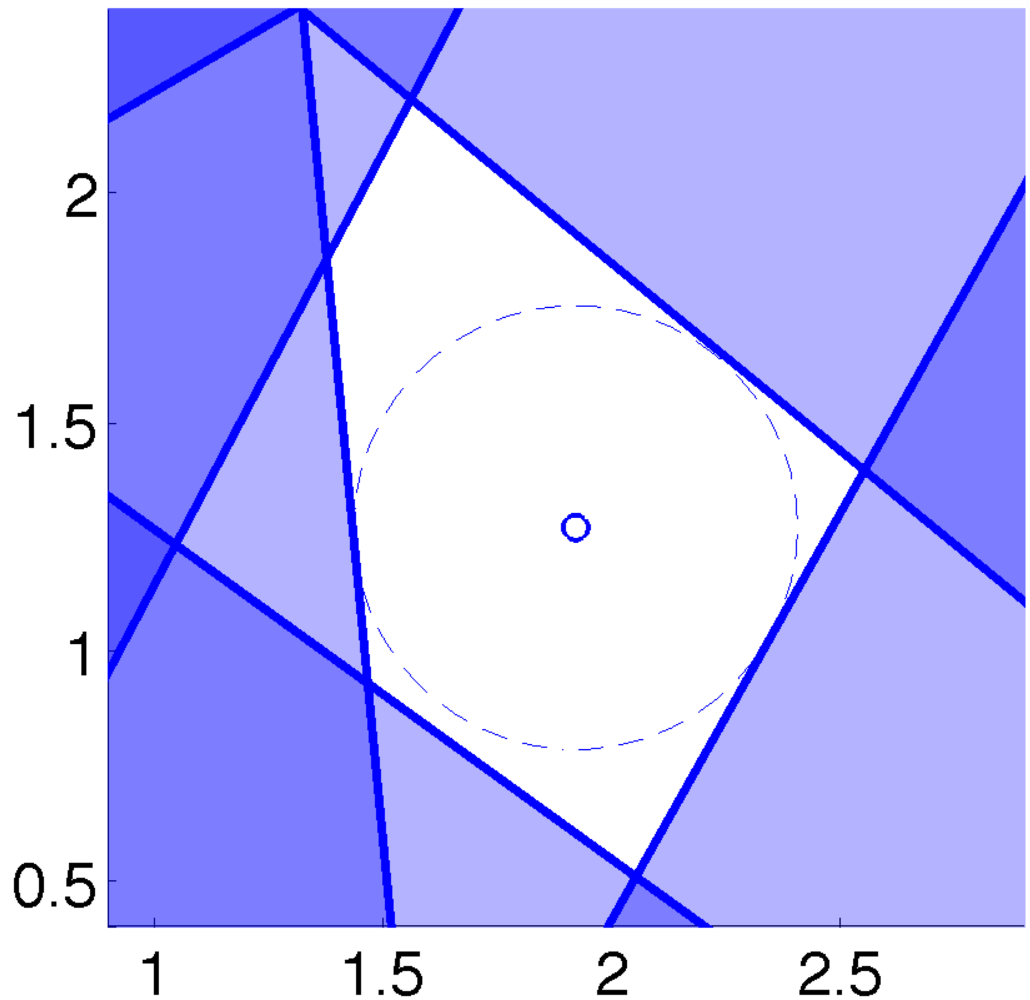
- $X = \{ x \mid Ax + b \geq 0 \}$
- Ball center:
  - $\max_x \min_i \text{dist}(x, \underline{a_i \cdot x + b_i = 0})$

    $a_i \cdot x + b_i$
  - if $||a_i|| = 1$

    $\max_{x,t} t \quad \text{st} \quad s = Ax + b \geq t1$
  - in general:

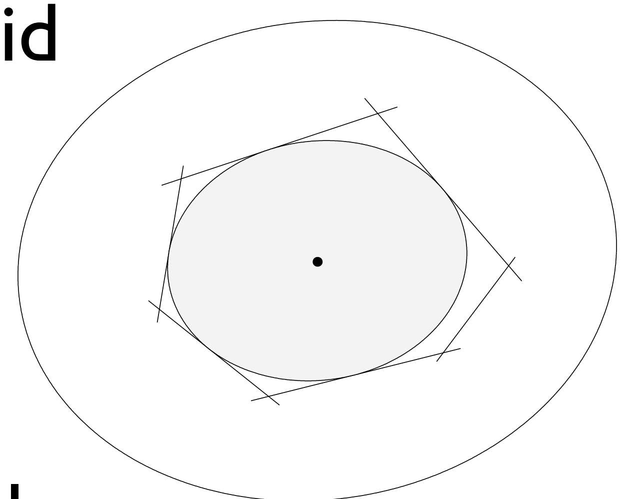
    $s_i / ||a_i|| \geq t \quad \forall i$





Ellipsoid center

aka max-volume inscribed ellipsoid

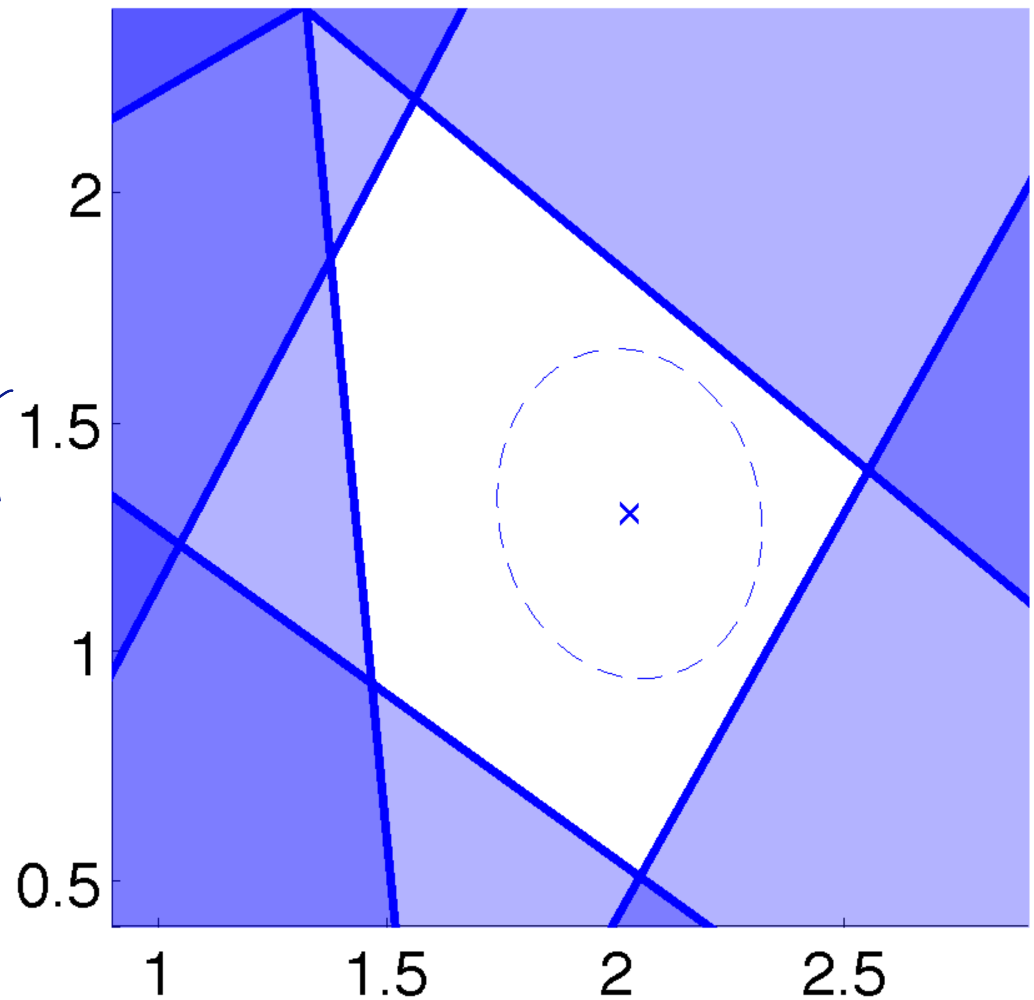


- Center d of largest inscribed ellipsoid
  - $E = \{ Bu + d \mid \|u\|_2 \leq 1 \}$
  - $\mathrm{vol}(E) \geq \mathrm{vol}(X)/n$ in R^n
- min log det B^{-1} s.t.
  - $a_i^T(Bu+d) + b_i \geq 0 \quad \forall i \quad \forall u$ with $\|u\| \leq 1$
  - $B \succcurlyeq 0$
- Convex optimization, but relatively expensive:
  - convex objective, semidefinite constraint
  - each (u, a_i, b_i) yields a linear constraint on B, d

Analytic center

- Let s = Ax + b
- Analytic center:
  - $\max_{x,s\geq 0} \prod_i s_i/\|a_i\|$
  - $\min_{x,s\geq 0} -\sum \ln s_i/\|a_i\|$

    $= -\sum \ln s_i + \cancel{\sum \ln \|a_i\|}$

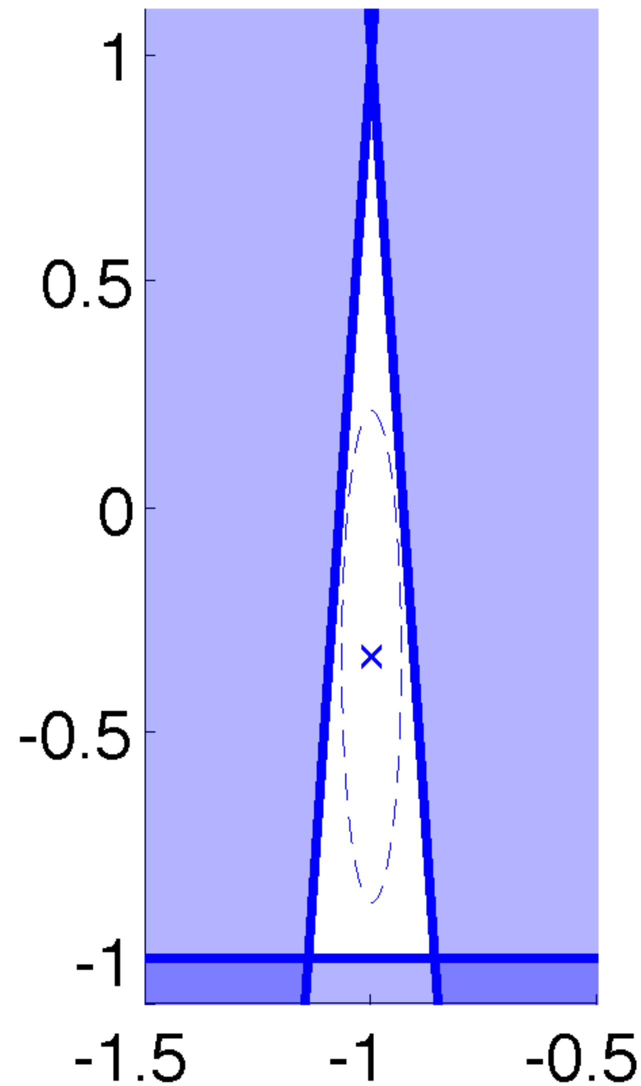


Bad conditioning? No problem.

$a_i^T x + b_i \geq 0 \quad \min -\sum \ln(a_i^T x + b_i)$

$- \sum_i \frac{1}{a_i \cdot x + b_i} a_i$

$y = Mx + q$



Newton for analytic center

$s = Ax + b \qquad S = \mathrm{diag}(s)$

- $f(x) = -\sum \ln(a_i^T x + b_i)$ $= -\sum_i$
  - $df/dx = -\sum a_i / (a_i^T x + b_i)$ $= -\sum_i \frac{1}{s_i} a_i = -A^T \frac{1}{s}$
  - $d^2f/df^2 =$ $\sum_i \frac{1}{(a_i^T x + b_i)^2} a_i a_i^T = A^T S^{-2} A$

Adding an objective

- Analytic center was for: find x st $Ax + b \geq 0$
- Now: min $c^T x$ st $Ax + b \geq 0$
- Same trick:
  - min $f_t(x) = c^T x - (1/t) \sum \ln(a_i^T x + b_i)$
    - parameter $t > 0$
  - central path =
    - $t \to 0$: analytic center $t \to \infty$: LP opt