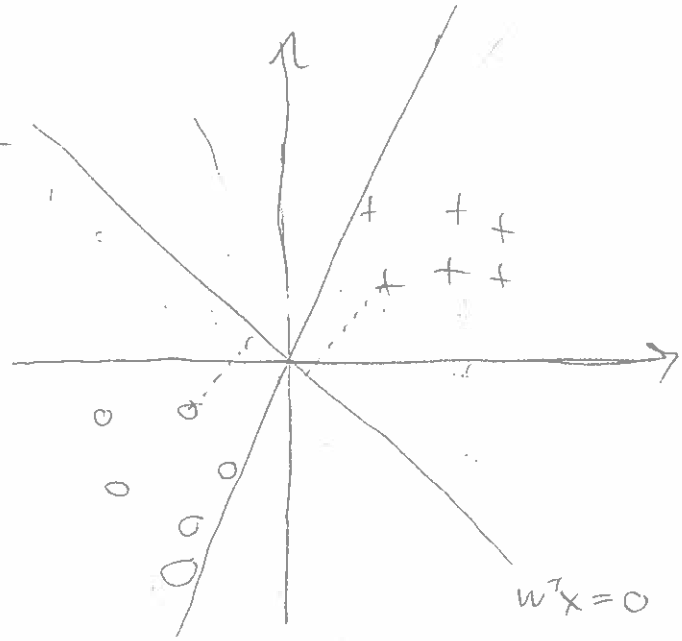
SVMs



Linear separator:

$w^Tx_i \geq 0$ if $y_i = 1$
$w^Tx_i < 0$ if $y_i = 0$
} "separates" training data

Decision boundary

$\{x : w^Tx = 0\}$

<u>Goal</u>: Find linear separator with biggest <u>margin</u>.

<u>margin</u> = distance from closest training example to decision boundary

---

(Note about illustrations:

General linear function: $h(x) = w^Tx + w_0$ — Decision boundary $h(x)=0$: $w^Tx = -w_0$

Simplification: $h(x) = w^Tx$ — $w^Tx = 0$

Implies decision boundary should go through origin in illustrations. Assume data is "centered")

In practice, either:

1) Add constant feature $(x_1, \ldots, x_d) \longmapsto (1, x_1, \ldots, x_d)$
2) Treat w_0 separately in math.

How to solve this problem?

objective function $J(w) = \|w\|^2$

$$(*) \quad \min_{w} \|w\|^2$$
$$\text{subject to } \begin{rcases} w^T x_i \geq 1 & \text{if } y_i = 1 \\ w^T x_i \leq -1 & \text{if } y_i = 0 \end{rcases} \text{"constraints"}$$

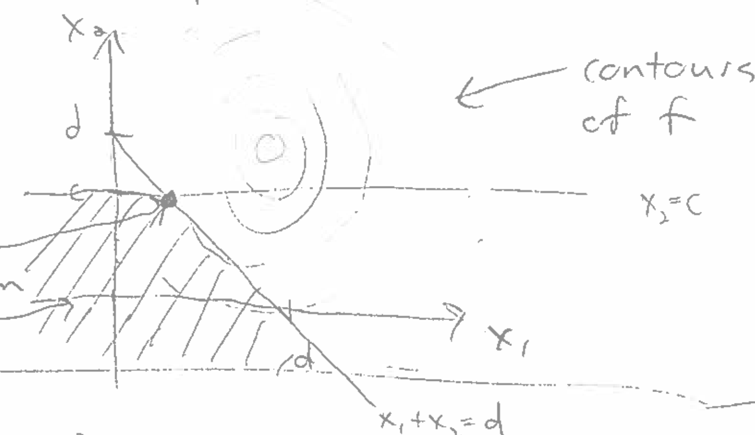
constrained optimization problem

Constrained optimization example (2d)

$$\min_{x_1, x_2} f(x_1, x_2)$$
$$\text{s.t. } x_2 \leq c$$
$$x_1 + x_2 \leq d$$

optimum

feasible region



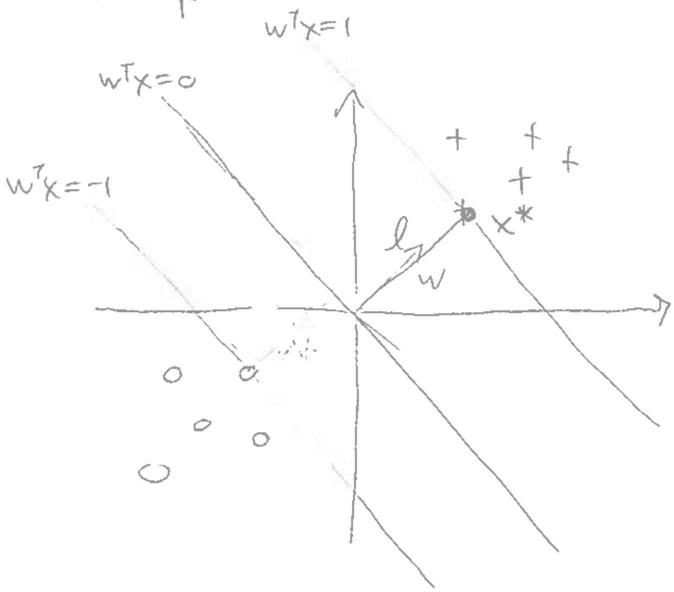
Geometric intepretation of $(*)$

Find w to minimize $\|w\|^2$ (slope!), such that

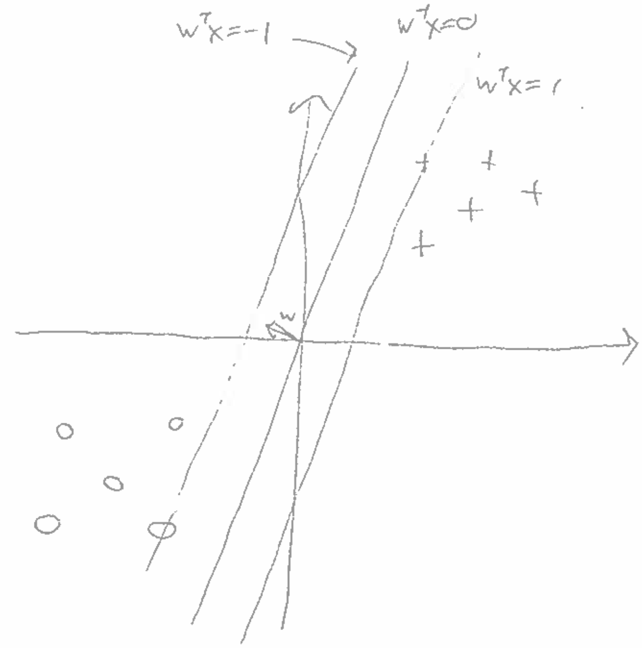
1) $h_w(x) \geq 1$: positive examples have "height" at least 1
2) $h_w(x) \leq 1$: negative examples have "height" at most -1

$\Rightarrow$ at least one contour line away from decision bdry.

Example



Good orientation

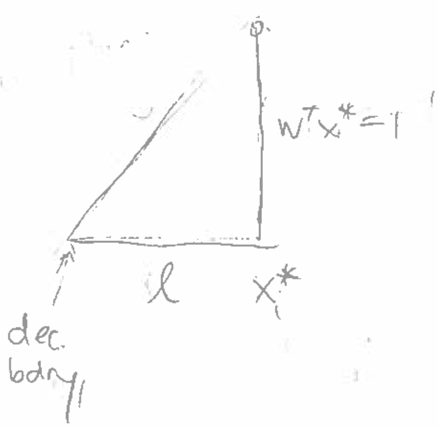


Bad orientation

MATLAB demo

Does (*) maximize margin?

- Let x^* be point closest to dec. bdry
- " ℓ be margin
- Claim: $w^T x^* = 1$ (otherwise make slope smaller)



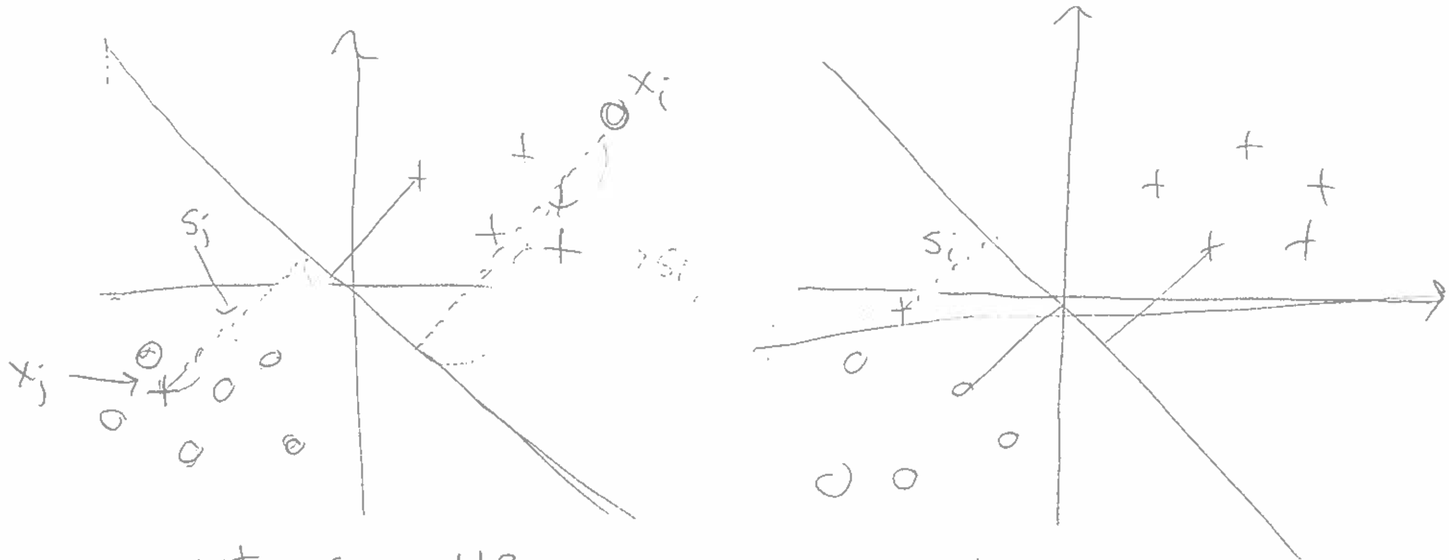
$$\frac{1}{\ell} = \text{slope} = \|w\|$$

$$\ell = \frac{1}{\|w\|} = \text{margin}$$

minimize $\|w\|^2 \Rightarrow$ maximize $\frac{1}{\|w\|} = \ell \Rightarrow$ maximize margin

Soft-margin SVMs

What if data is not linearly separable?
(e.g. noise)



not separable $\longleftrightarrow$ barely separable

- Idea: allow example to be on wrong side, for a cost
- Penalty for example x_i:

$$s_i = \begin{cases} 0 & \text{if } x_i \text{ on correct side} \\ \text{distance from bdry} & x_i \text{ on wrong side} \end{cases}$$

constant

$$\boxed{\begin{aligned} &\min_w \|w\|^2 + C \sum_{i=1}^{N} s_i \\ &\text{s.t. } w^T \vec{x}_i \geq 1 - s_i \quad y_i = 1 \\ &\qquad w^T \vec{x}_i \leq 1 + s_i \quad y_i = 0 \\ &\qquad s_i \geq 0 \end{aligned}}$$ (**)

misclassification penalty

Comparison: soft-margin SVM vs. regularized logistic regression

Let's massage (**):

① Multiply objective by $\lambda = \frac{1}{C}$

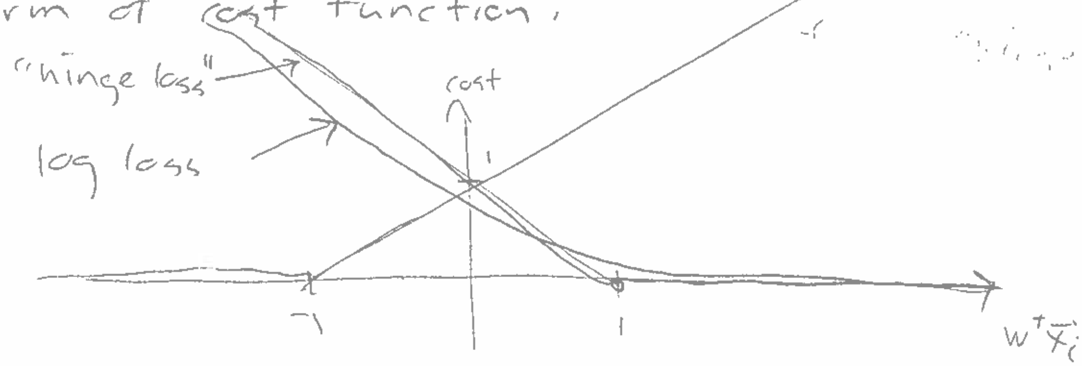
$$\Rightarrow \min_{w} \; \lambda \|w\|^2 + \sum_{i=1}^{N} s_i$$

② Interpret s_i as cost function

$$s_i = \text{cost}(w^T \vec{x}_i, y_i)$$

$$\Rightarrow \boxed{\min_{w} \; \lambda \|w\|^2 + \sum_{i=1}^{N} \text{cost}(w^T \vec{x}_i, y_i)}$$

- Identical form to reg. logistic regression
- Form of cost function?



$y_i = 1$ $\quad$ $\text{cost}(w^T x_i, 1) = \begin{cases} 0 & \text{if } w^T x_i \geq 1 \\ 1 - w^T x_i & \text{if } w^T x_i < 1 \end{cases}$

$y_i = 0$ $\quad$ similar