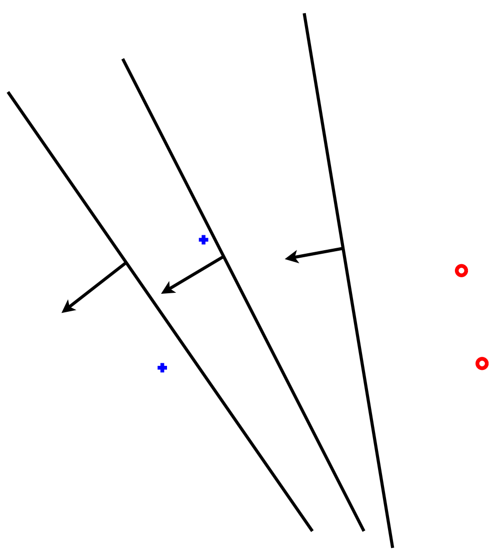
Machine Learning
Lecture 3

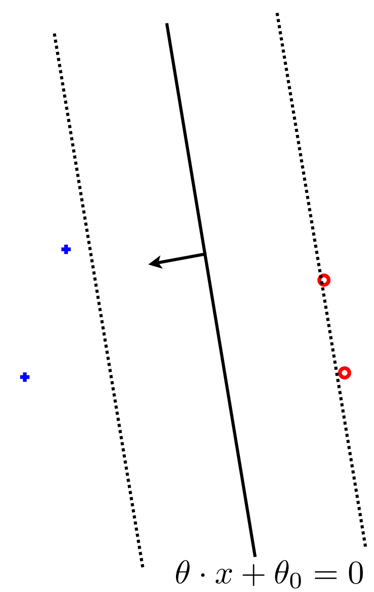
Outline

- Linear, large margin classification
  - margin, hinge loss, regularization
- Learning as an optimization problem

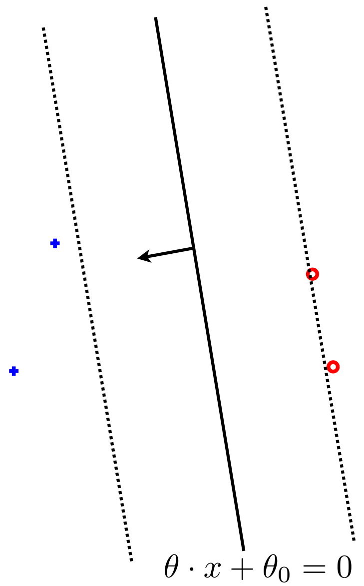
Linear classification



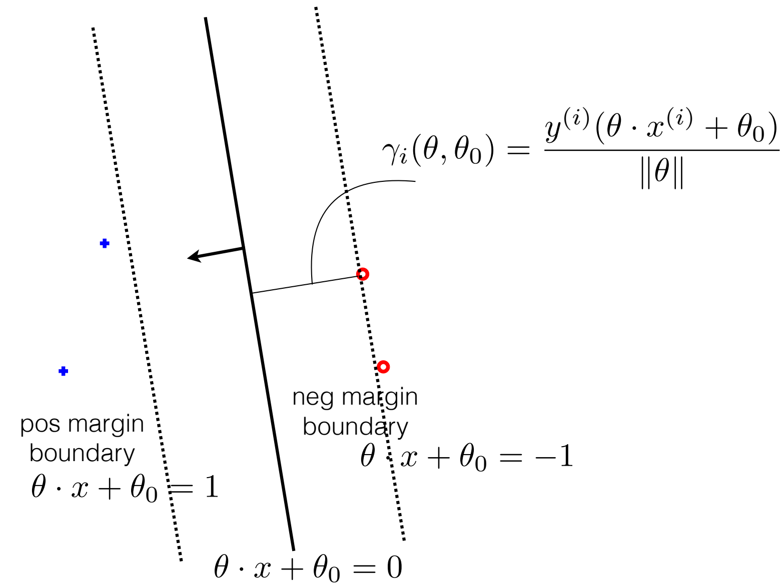
Learning as optimization



Learning as optimization



Linear classification, margin



Large margin as optimization

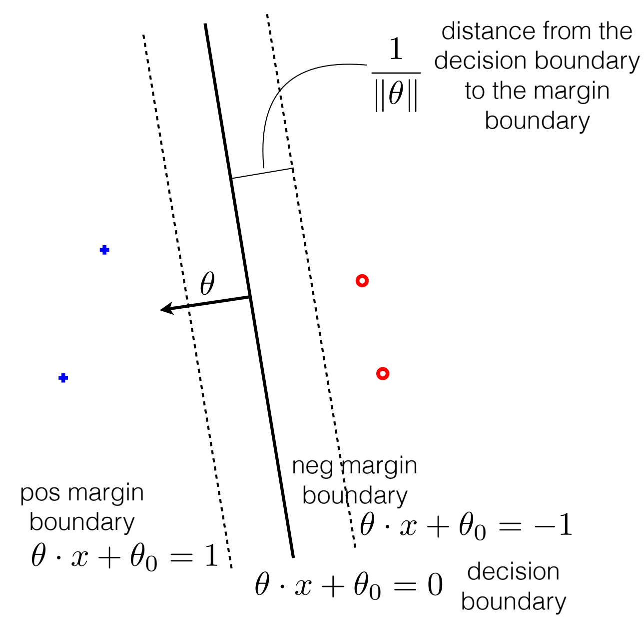
- Hinge loss

$$\text{Loss}_h\big(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\big) =$$

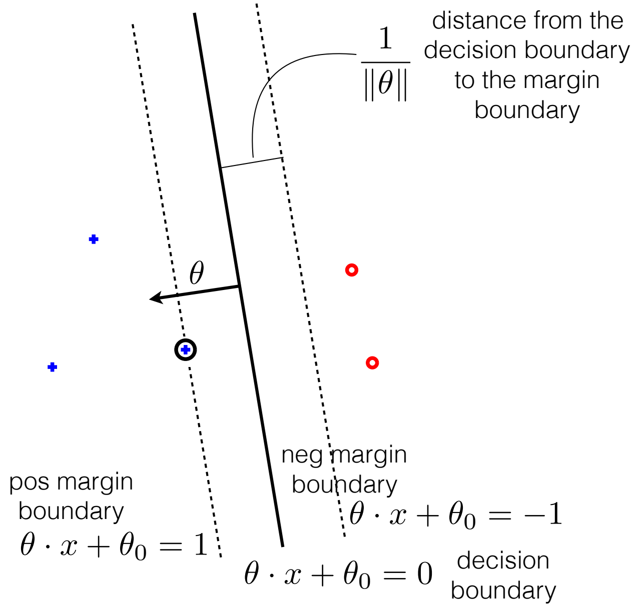
- Regularization: towards max margin

- The objective

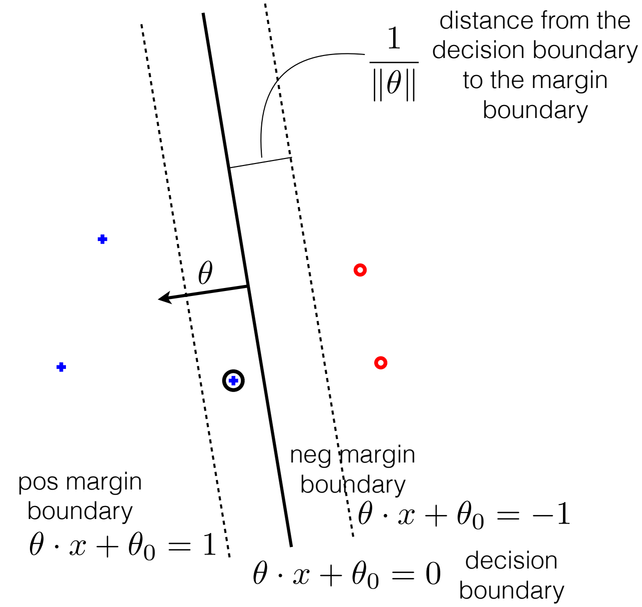
$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h\big(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\big) + \frac{\lambda}{2}\|\theta\|^2$$



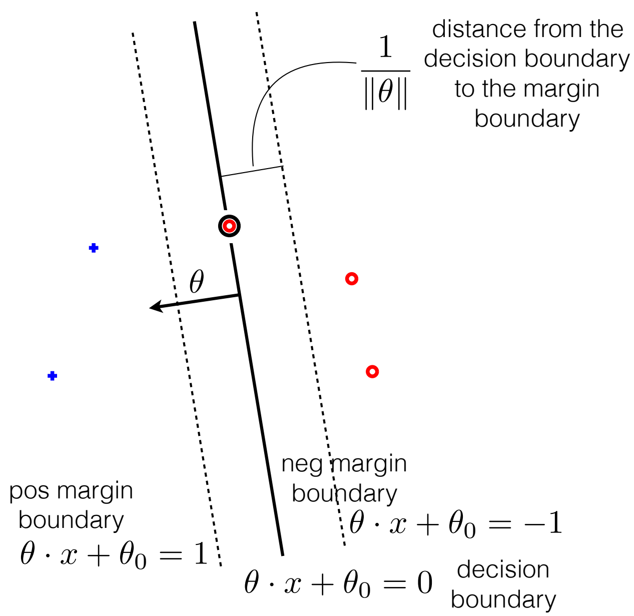
$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h\big(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\big) + \frac{\lambda}{2}\|\theta\|^2$$



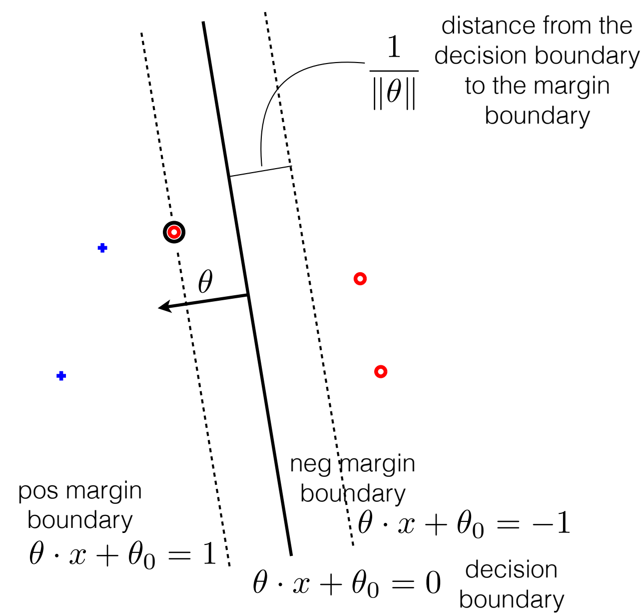
$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h\big(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\big) + \frac{\lambda}{2} \|\theta\|^2$$



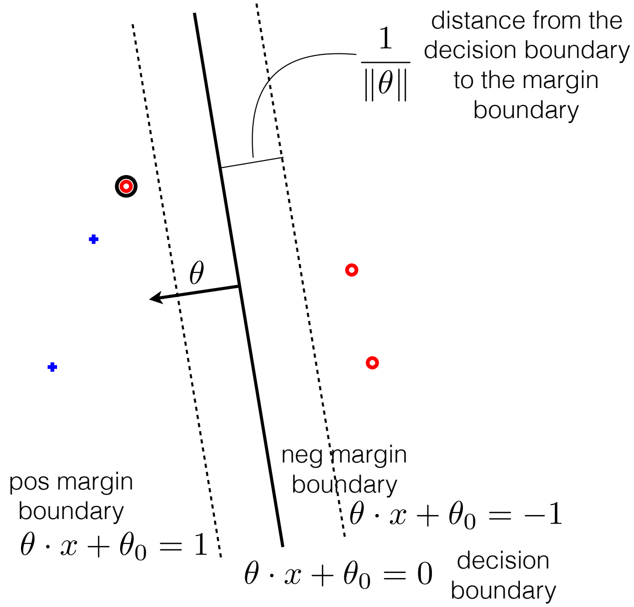
$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h\big(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\big) + \frac{\lambda}{2} \|\theta\|^2$$



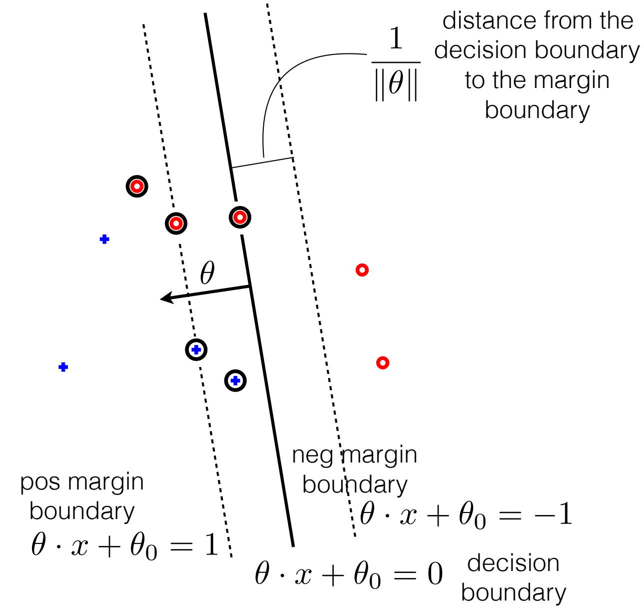
$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h\big(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\big) + \frac{\lambda}{2} \|\theta\|^2$$



$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h\big(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\big) + \frac{\lambda}{2} \|\theta\|^2$$



$$J(\theta, \theta_0) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Loss}_h\big(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\big) + \frac{\lambda}{2}\|\theta\|^2$$



$$J(\theta, \theta_0) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Loss}_h\big(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\big) + \frac{\lambda}{2}\|\theta\|^2$$

CSAIL

Things to know

- General optimization formulation of learning

  objective function = average loss + regularization

- Large margin linear classification as optimization
  - margin boundaries, hinge loss, regularization

$$J(\theta, \theta_0) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Loss}_h\big(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\big) + \frac{\lambda}{2}\|\theta\|^2$$