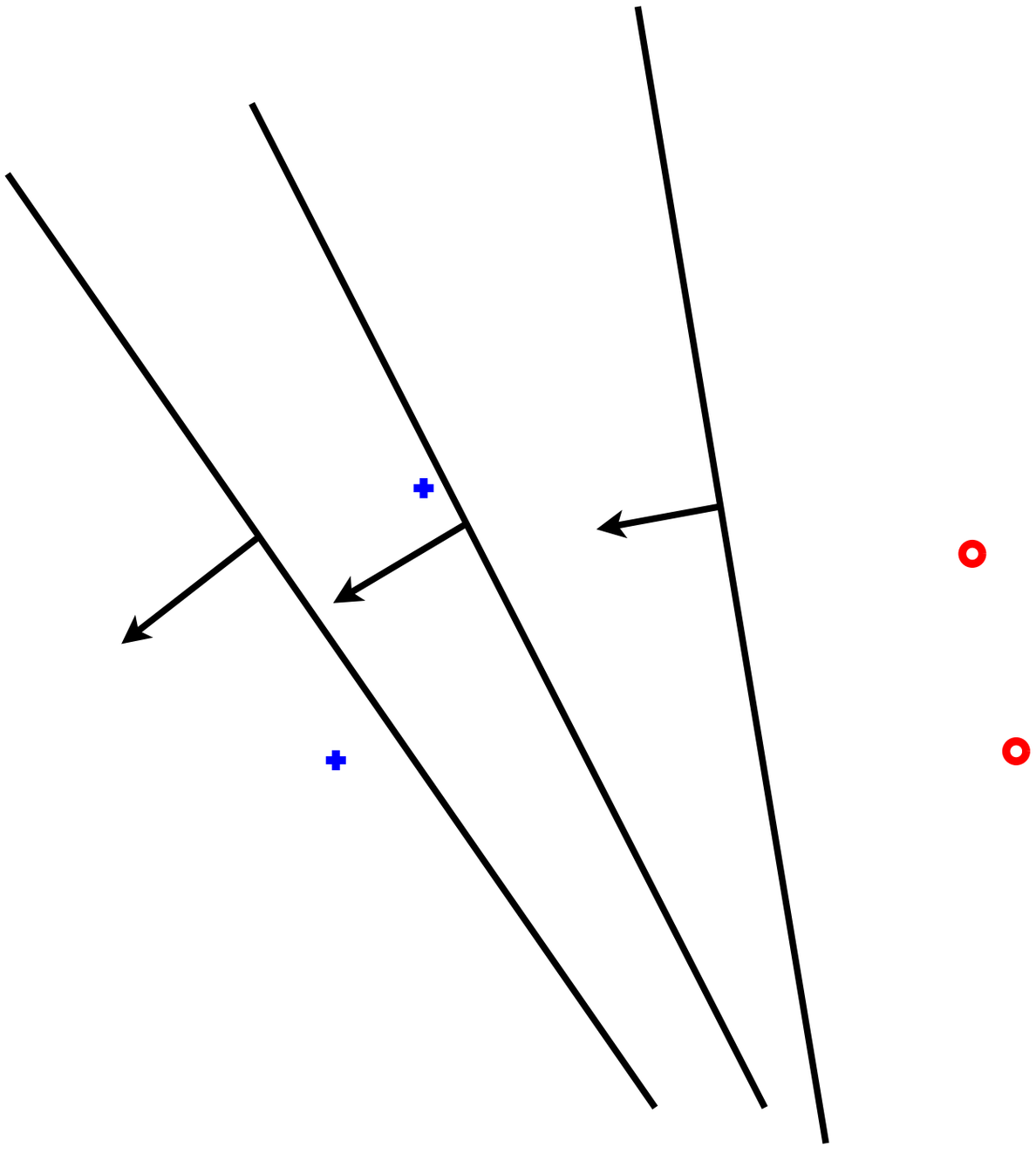
Machine Learning
Lecture 3

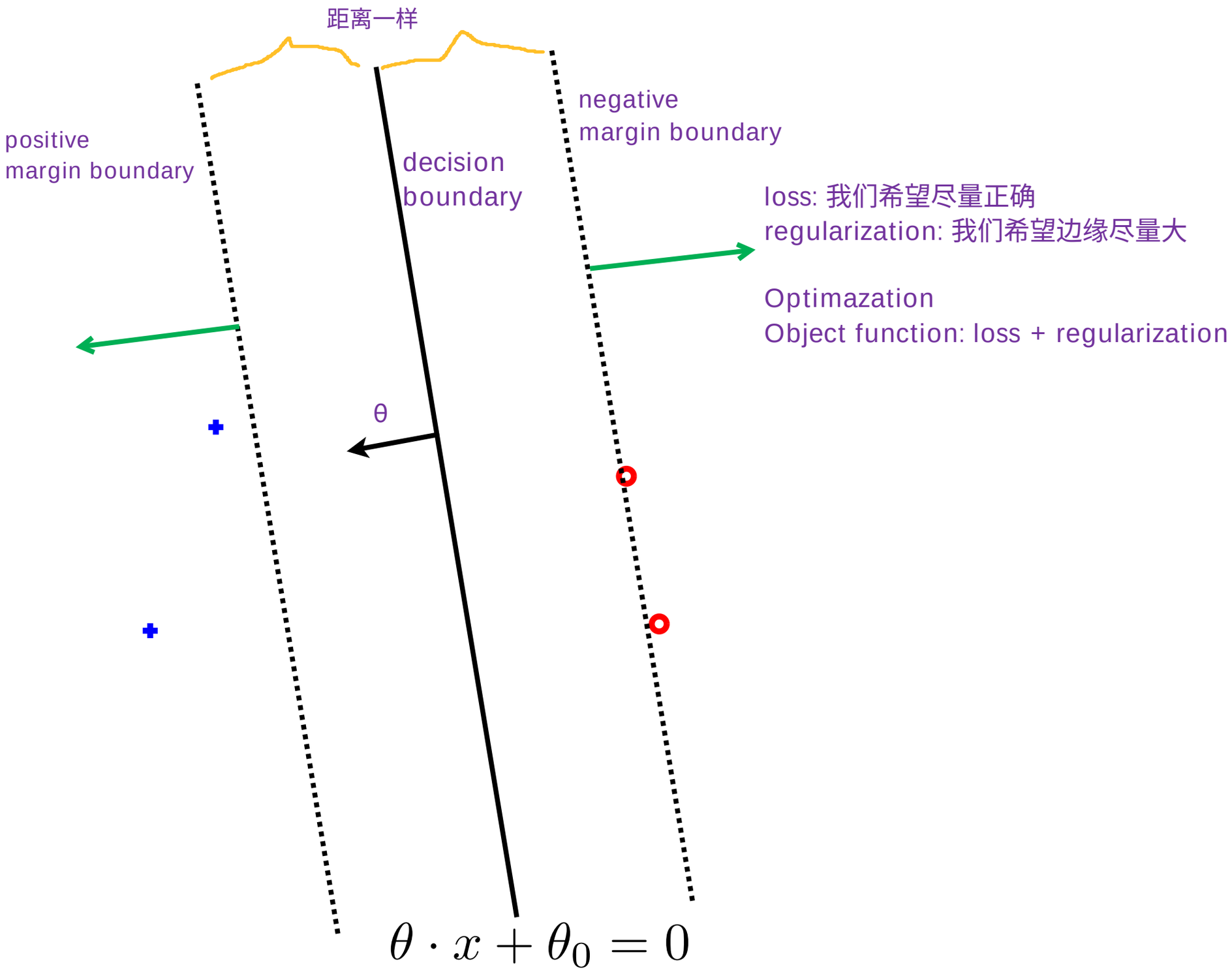
Outline

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

Linear classification



Learning as optimization

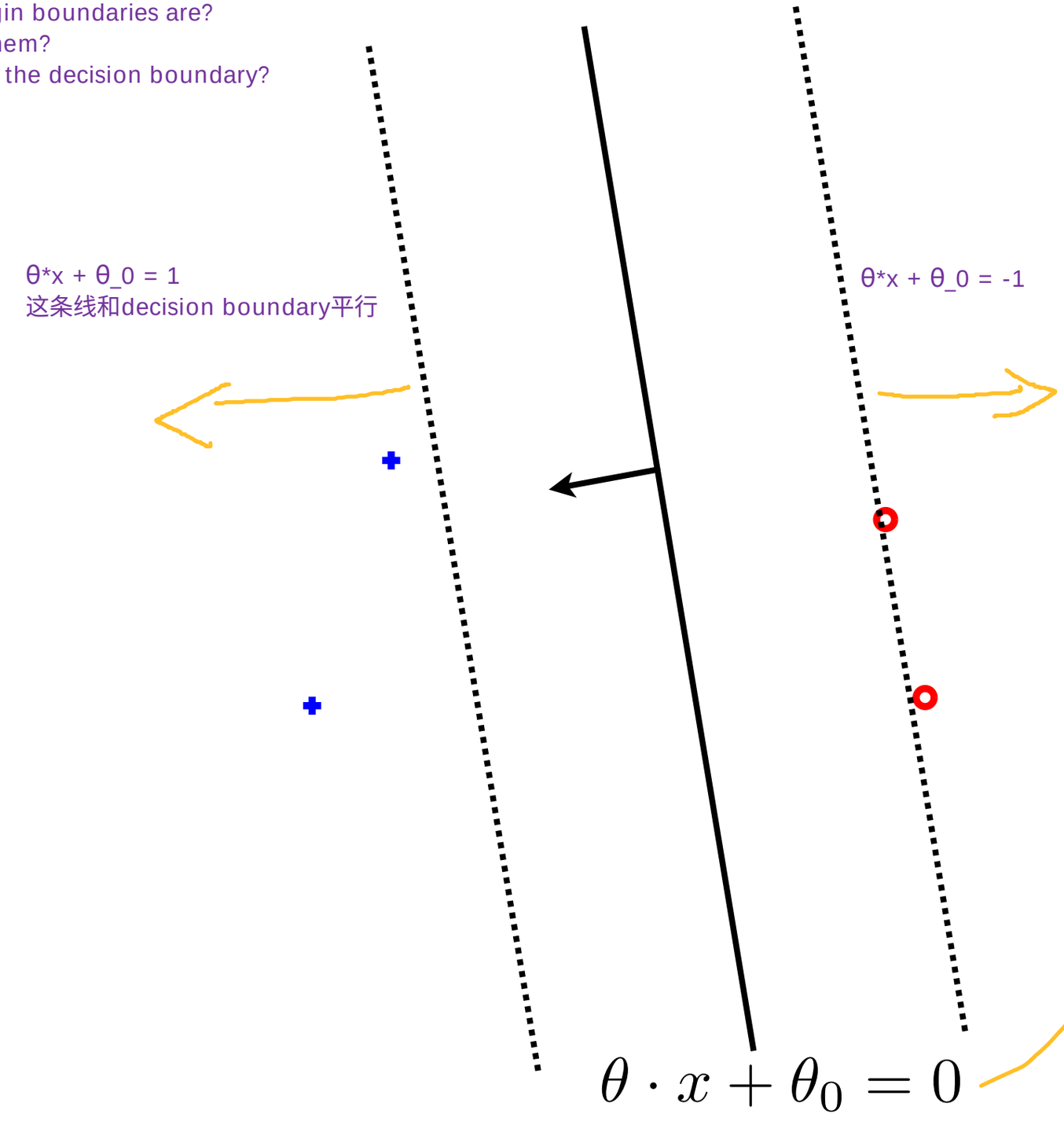


loss: 我们希望尽量正确
regularization: 我们希望边缘尽量大

Optimazation
Object function: loss + regularization

Learning as optimization

What exactly the margin boundaries are?
How we can control them?
How far they are from the decision boundary?

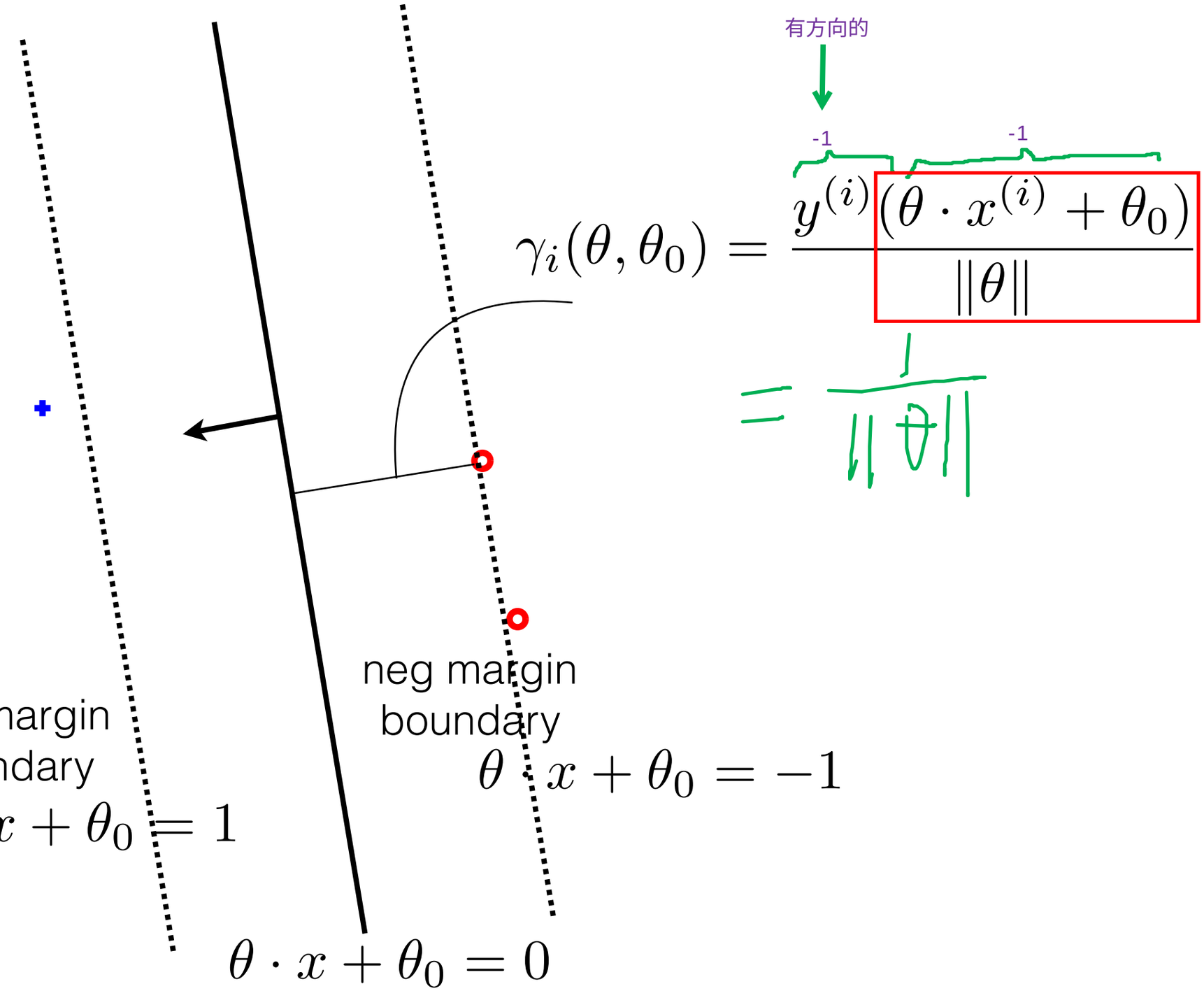


θ和θ_0定义的decision boundary的位置方向
norm(θ)定义了margin的长度

$$\frac{\theta}{\|\theta\|} x + \frac{\theta}{\|\theta\|} = 0$$

相等

Linear classification, margin



Large margin as optimization

- Hinge loss

agreement: z

$$\text{Loss}_h\left(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\right) = \begin{cases} 0 \text{ if } z >= 1 \\ 1\text{-}z \text{ if } z < 1 \text{(z越小，penalize的越厉害)} \end{cases}$$

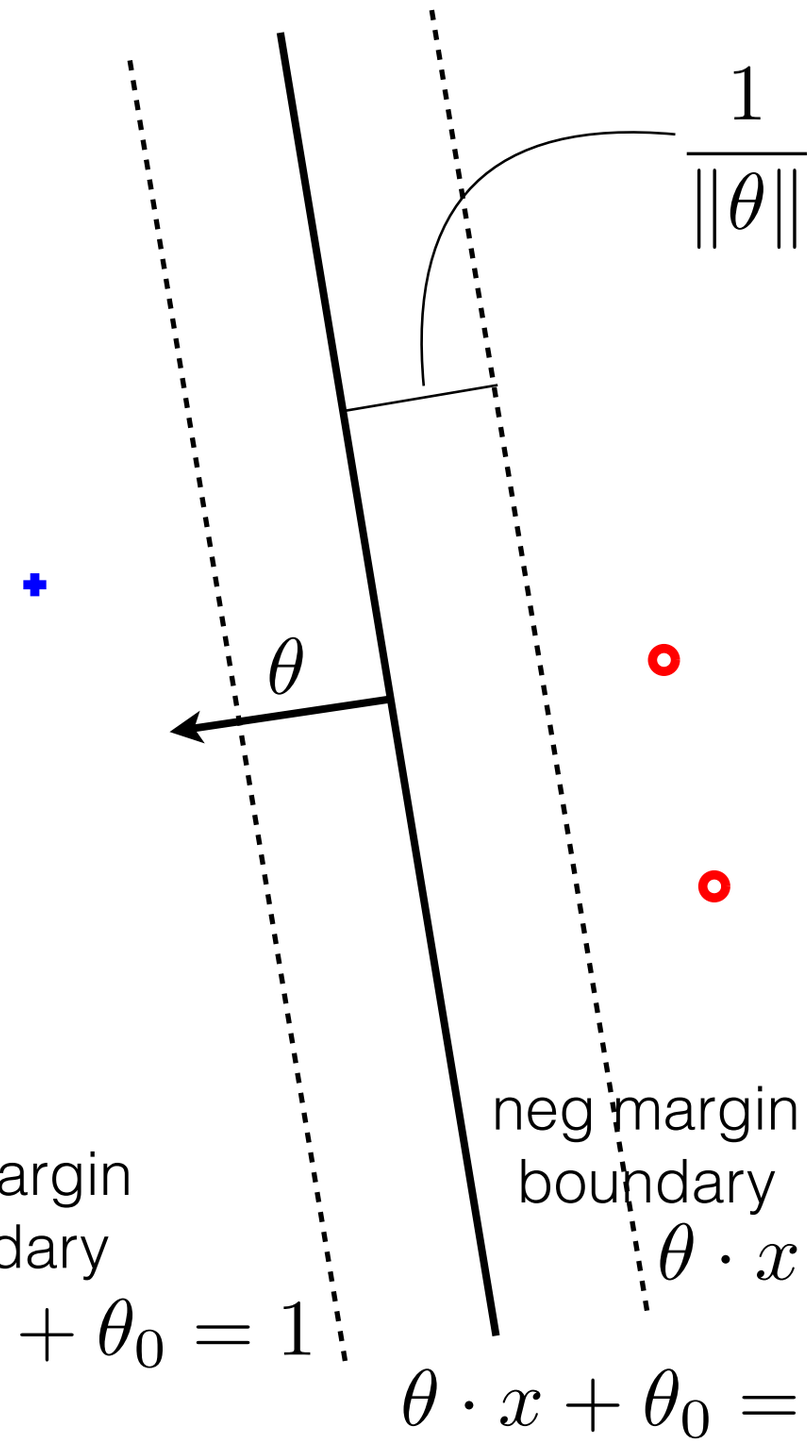
- Regularization: towards max margin

max 1/norm(θ) -> min norm(θ) -> min 1/2*norm(θ)^2

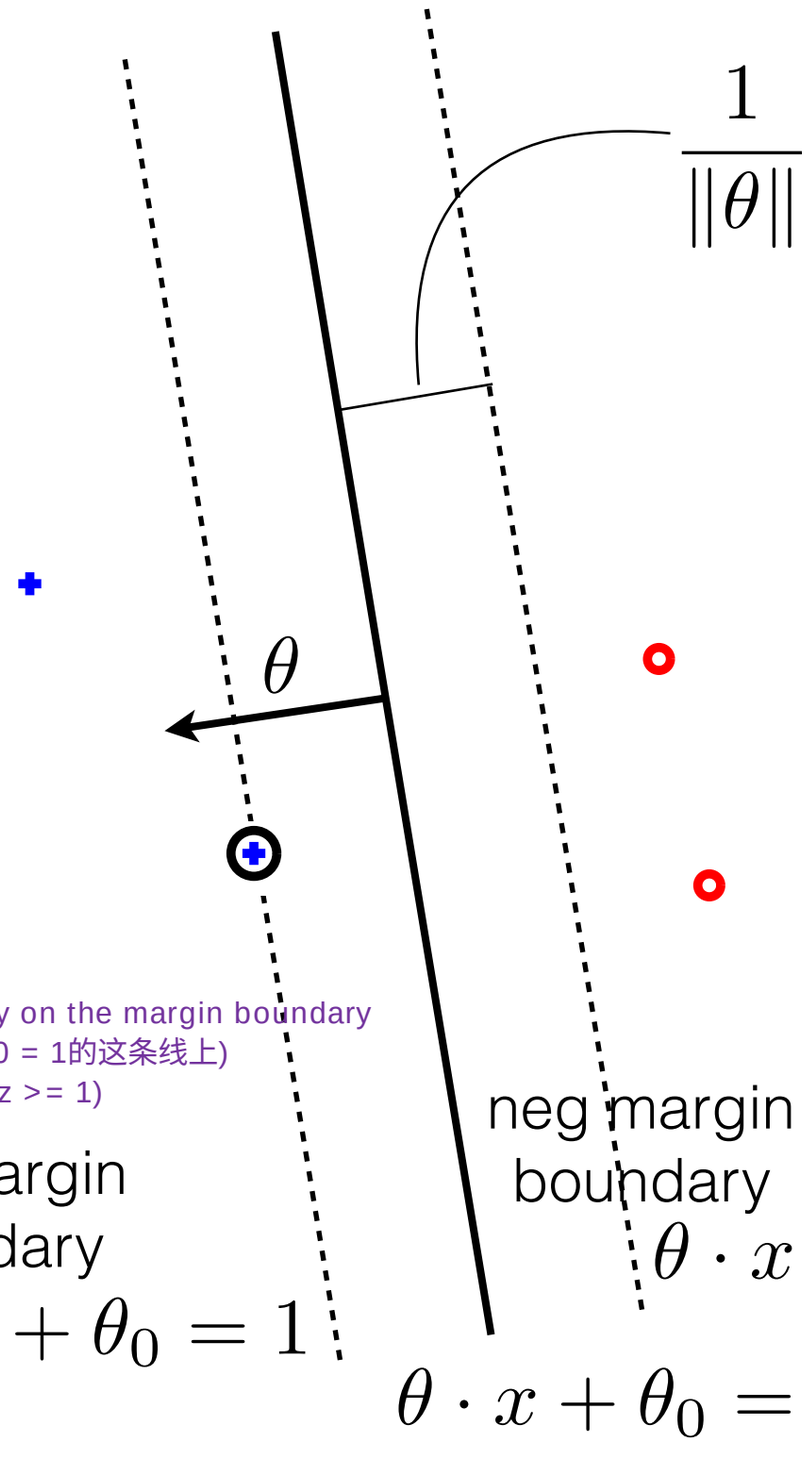
- The objective

regularization parameter >0

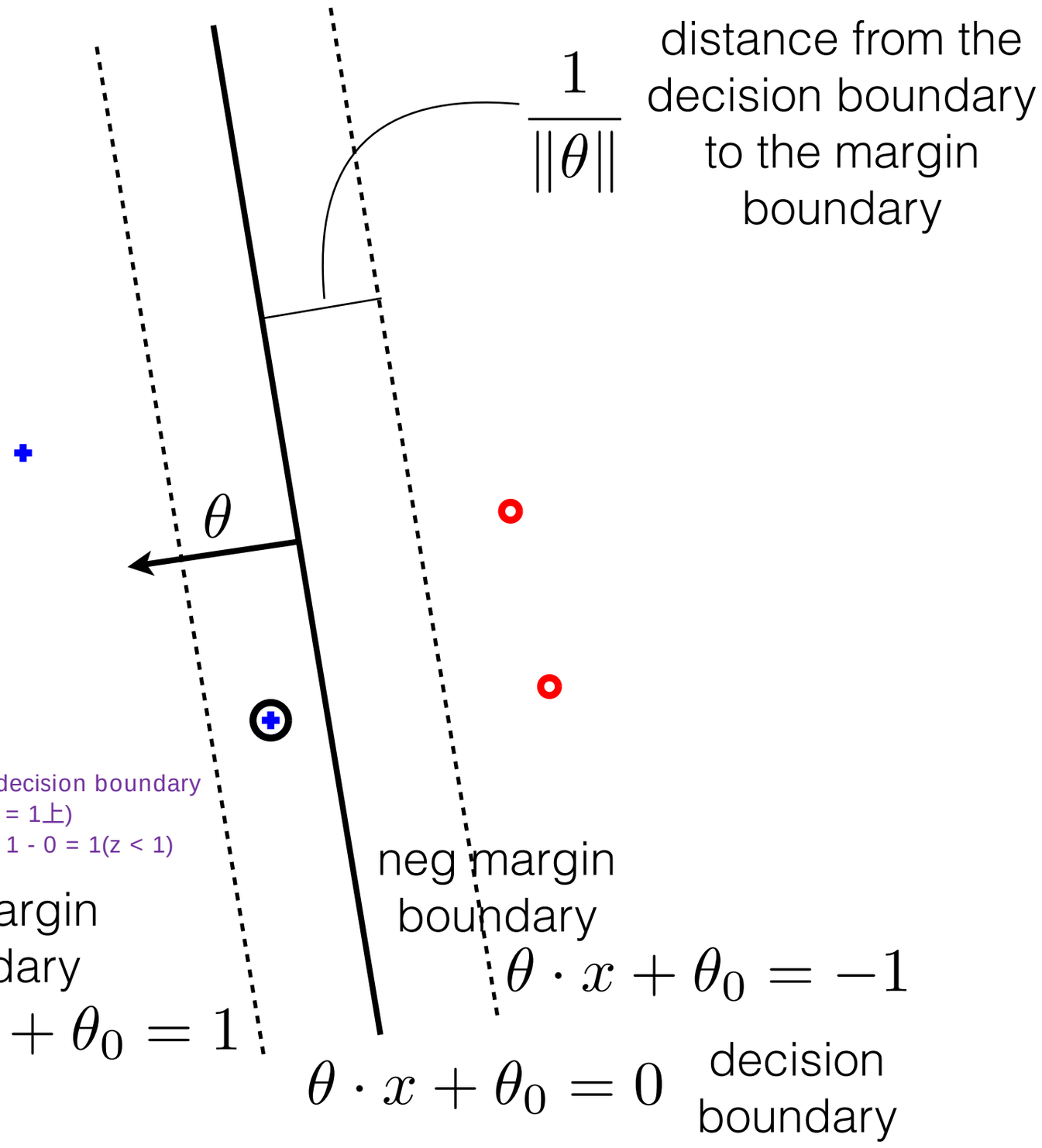
$$J(\theta, \theta_0) = \frac{1}{n}\sum_{i=1}^{n} \text{Loss}_h\left(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\right) + \frac{\lambda}{2}\|\theta\|^2$$



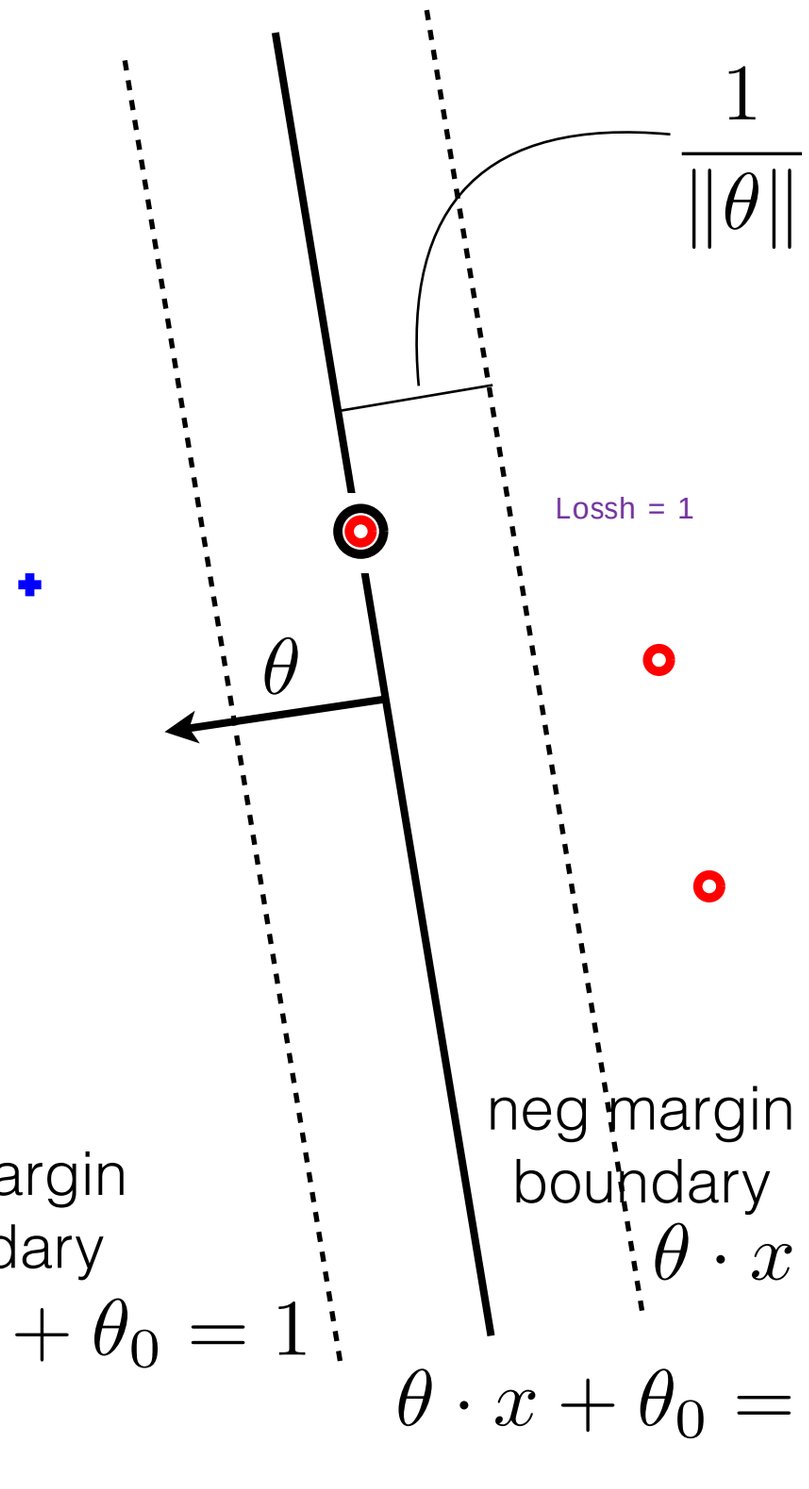
$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h\left(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\right) + \frac{\lambda}{2}\|\theta\|^2$$



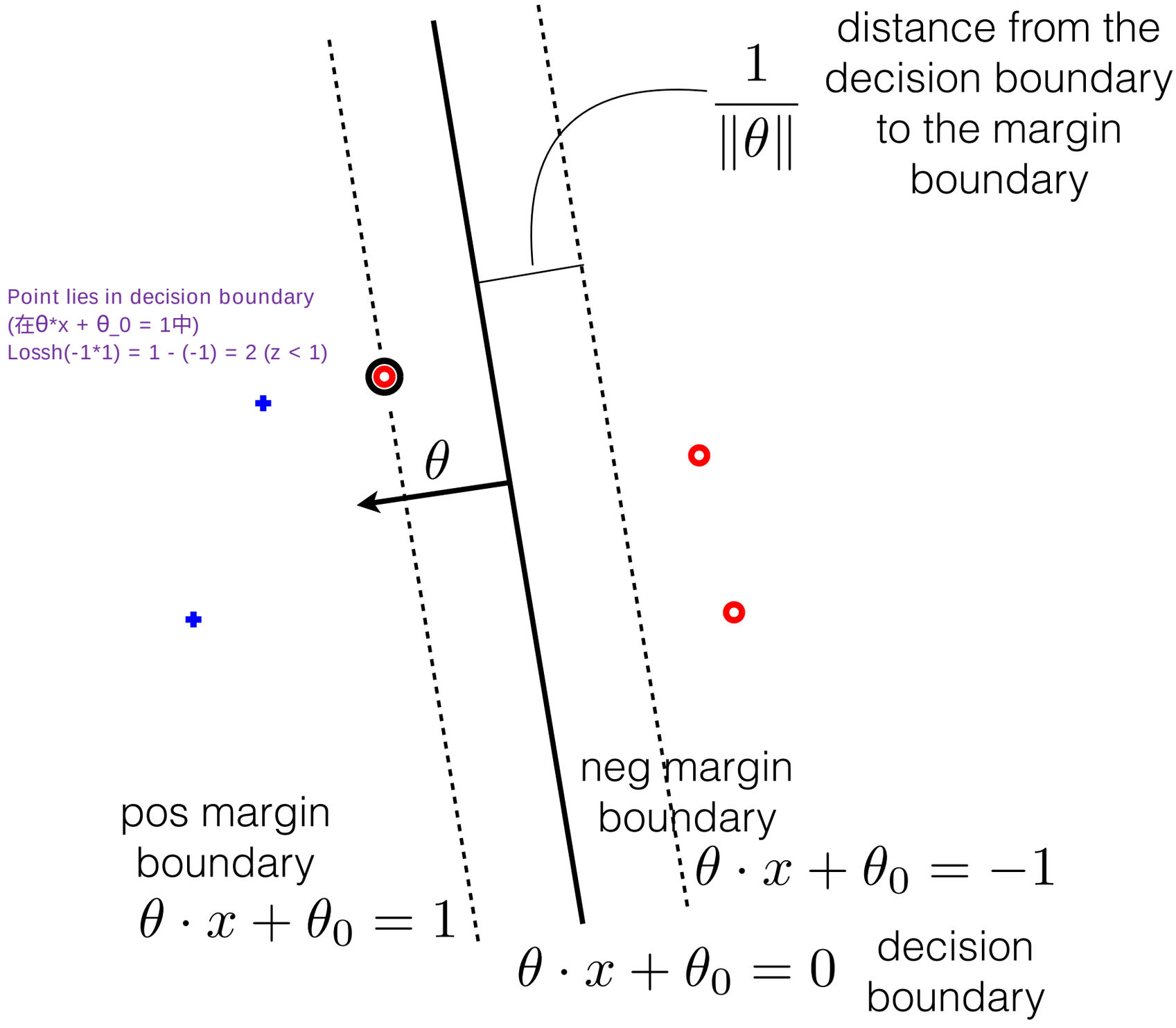
$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h\left(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\right) + \frac{\lambda}{2}\|\theta\|^2$$



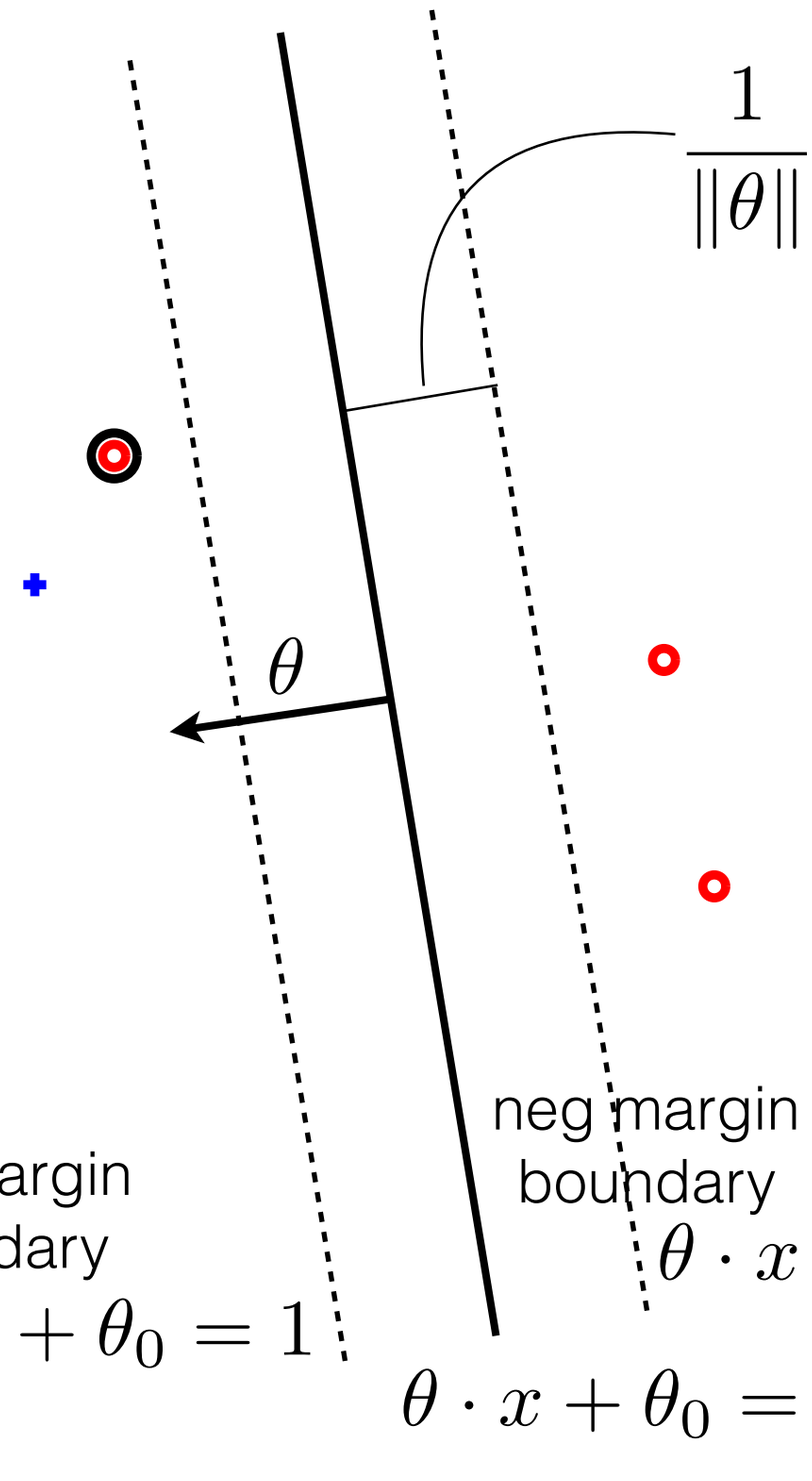
$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h\left(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\right) + \frac{\lambda}{2}\|\theta\|^2$$



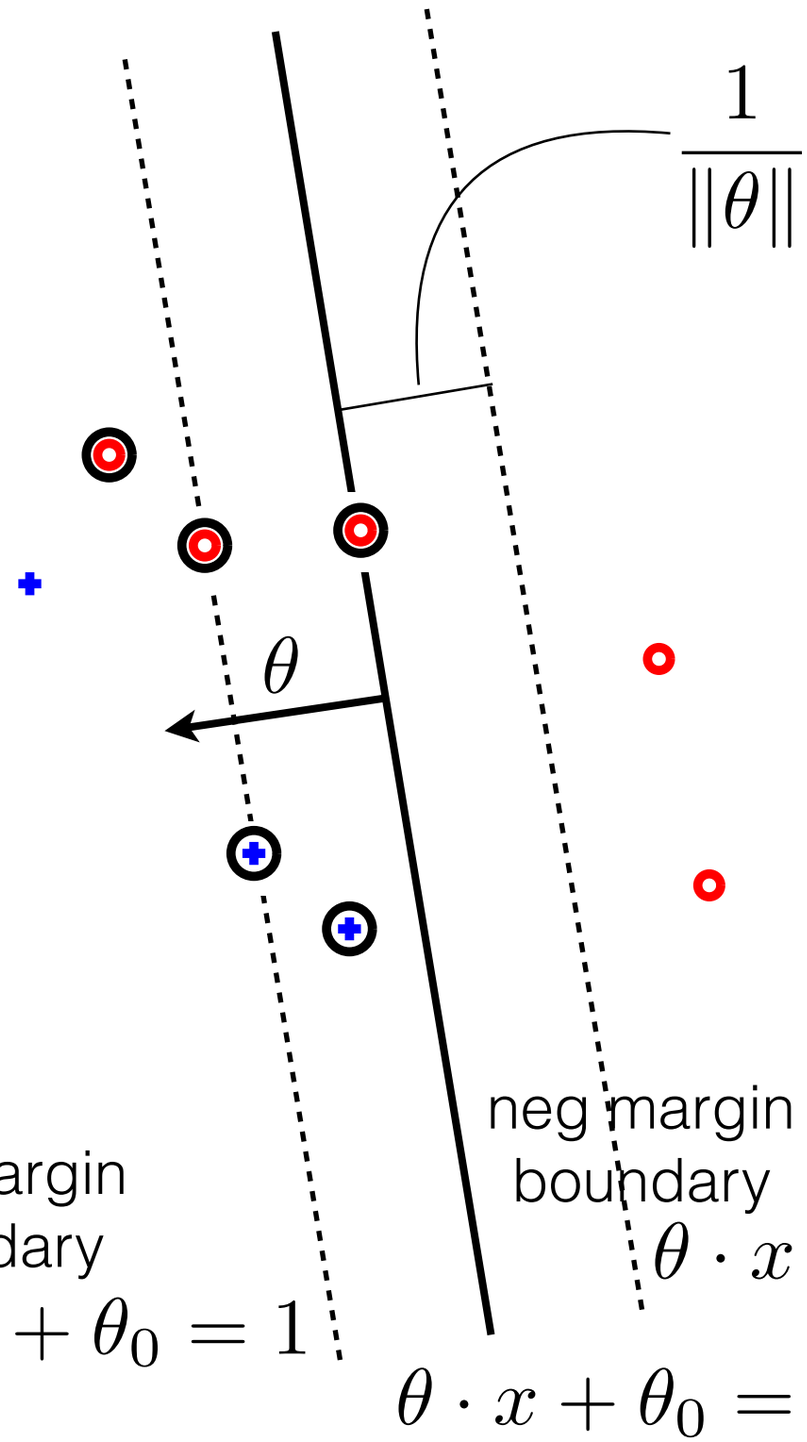
$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h\left(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\right) + \frac{\lambda}{2}\|\theta\|^2$$



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



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



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

Things to know

- General optimization formulation of learning

  objective function = average loss + regularization

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

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