Machine Learning
Lecture 3

Outline

CSAIL
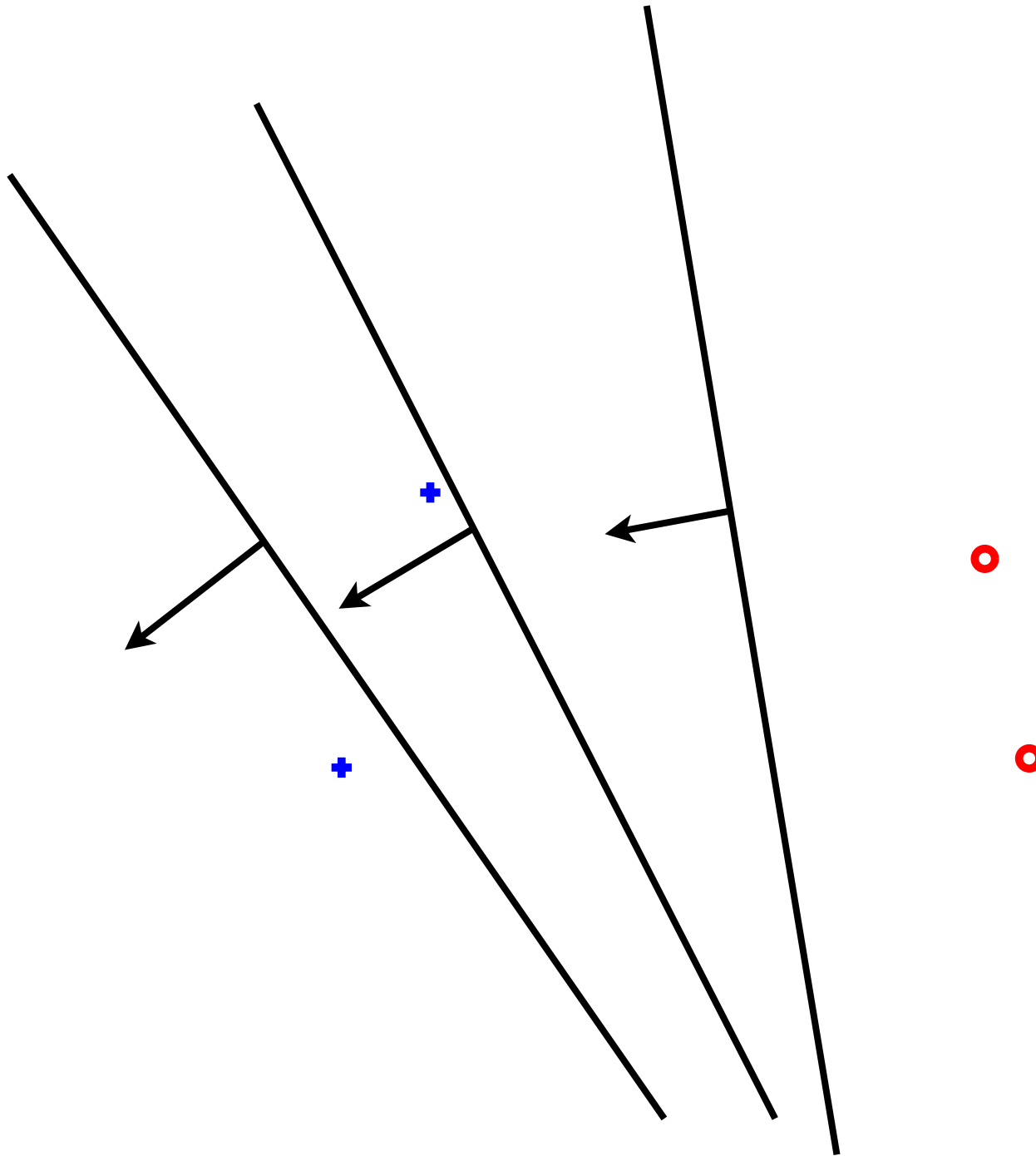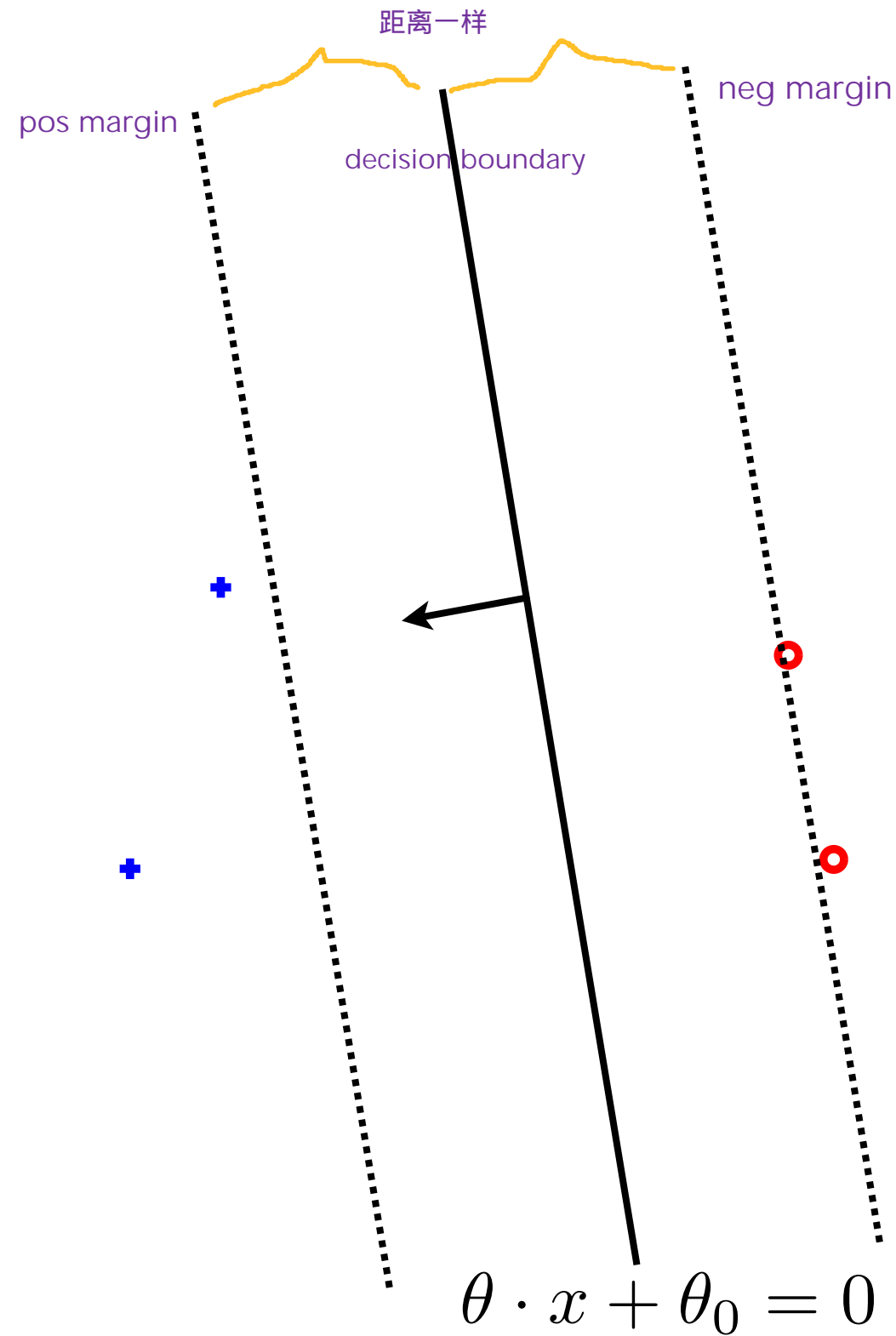
‣ Linear, large margin classification
  - margin, hinge loss, regularization

‣ Learning as an optimization problem

Linear classification

Learning as optimization

pos margin

neg margin

decision boundary

loss:
regularization:

Optimazation
Object function: loss + regularization

$$\theta \cdot x + \theta_0 = 0$$

Learning as optimization

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

$\theta^*x + \theta\_0 = 1$
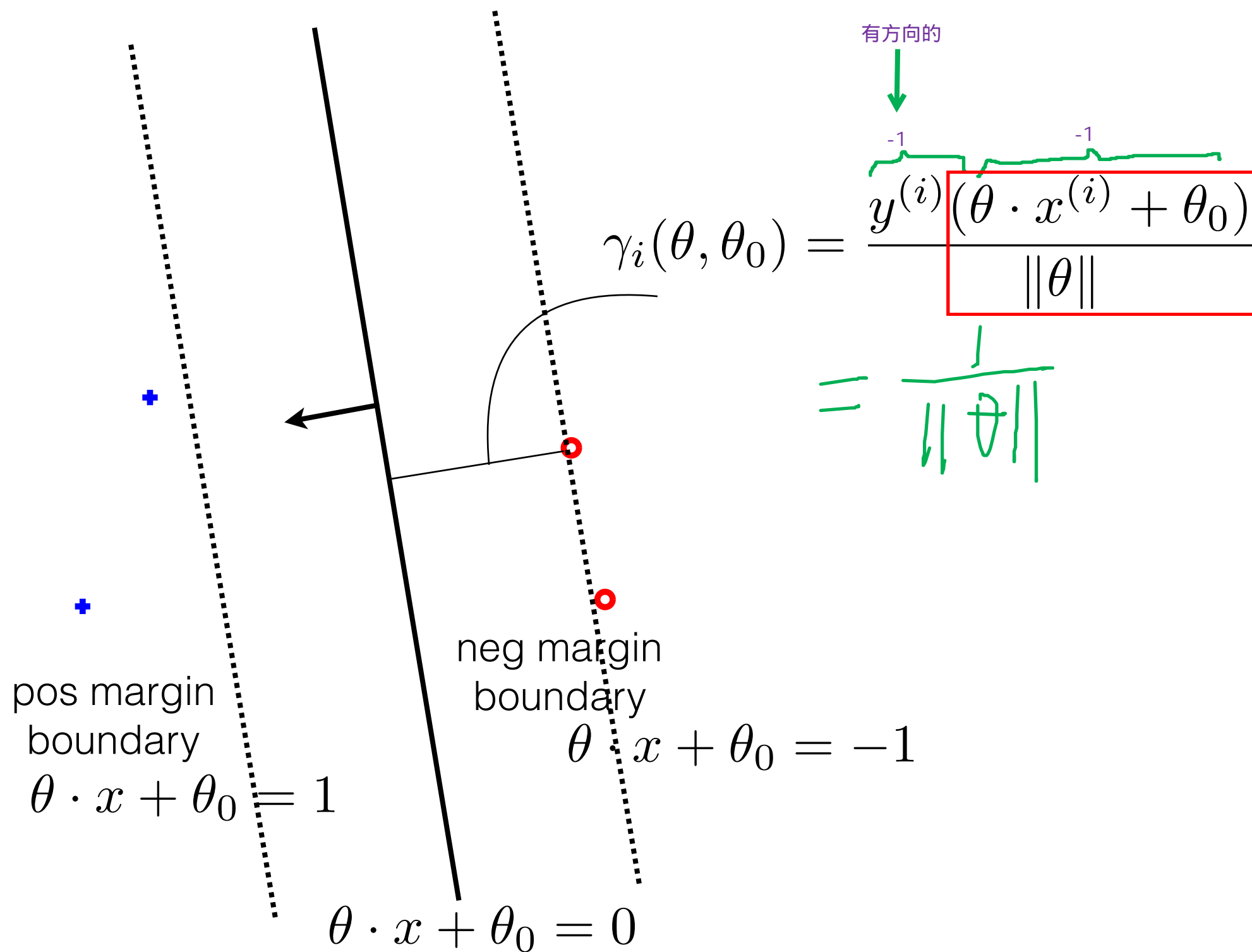decision boundary

$\theta^*x + \theta\_0 = -1$

$\theta \quad \theta\_0$
norm(θ)

decision boundary
margin

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

$$\theta \cdot x + \theta_0 = 0$$

Linear classification, margin

$$\gamma_i(\theta, \theta_0) = \frac{y^{(i)}(\theta \cdot x^{(i)} + \theta_0)}{\|\theta\|}$$

-1   -1

$$= \frac{1}{\|\theta\|}$$

pos margin
boundary
$$\theta \cdot x + \theta_0 = 1$$

neg margin
boundary
$$\theta \cdot x + \theta_0 = -1$$

$$\theta \cdot x + \theta_0 = 0$$

Large margin as optimization

‣ Hinge loss

agreement: z

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

‣ Regularization: towards max margin

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

‣ The objective

regularization parameter
>0

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

$$\frac{1}{\|\theta\|}$$ distance from the decision boundary to the margin boundary

θ

neg margin boundary

pos margin boundary

$$\theta \cdot x + \theta_0 = -1$$

$$\theta \cdot x + \theta_0 = 1$$

$$\theta \cdot x + \theta_0 = 0$$

decision boundary

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

distance from the decision boundary to the margin boundary

$$\frac{1}{\|\theta\|}$$

θ

Point lies exactly on the margin boundary
(θ*x + θ_0 = 1   )
Lossh(1*1) = 0 (z >= 1)

pos margin boundary

neg margin boundary

$$\theta \cdot x + \theta_0 = 1$$

$$\theta \cdot x + \theta_0 = 1$$

$$\theta \cdot x + \theta_0 = 0$$

decision boundary

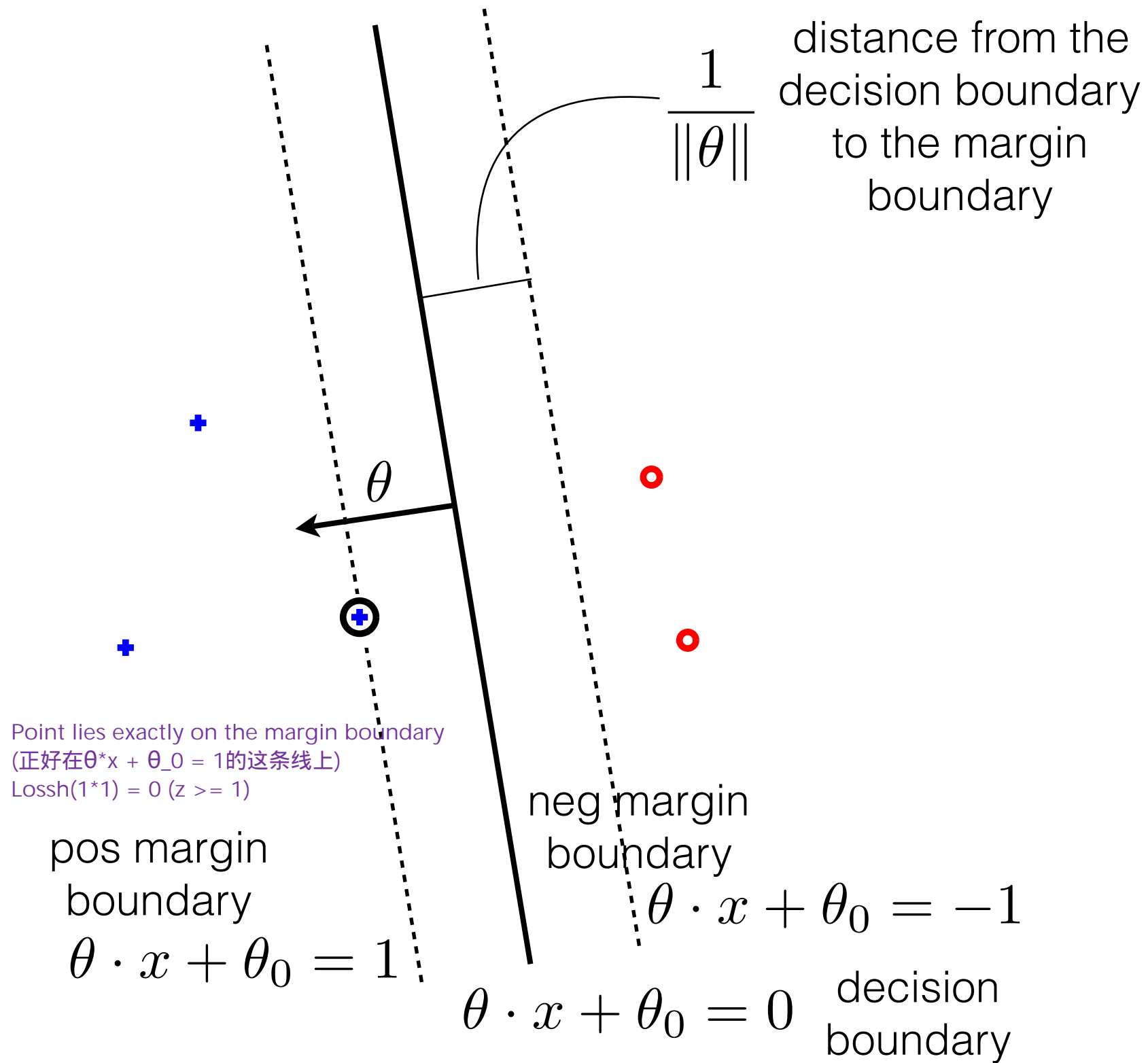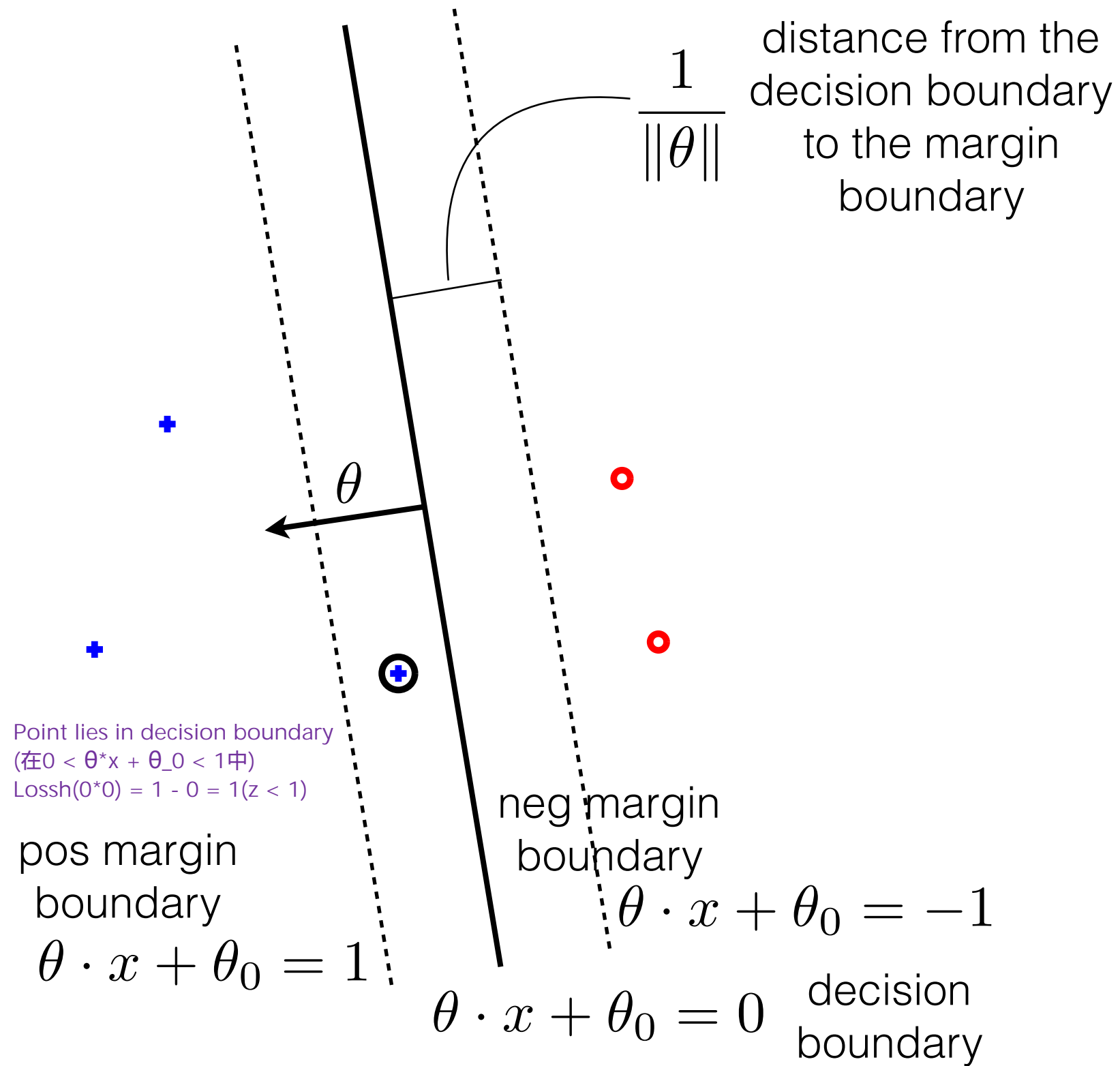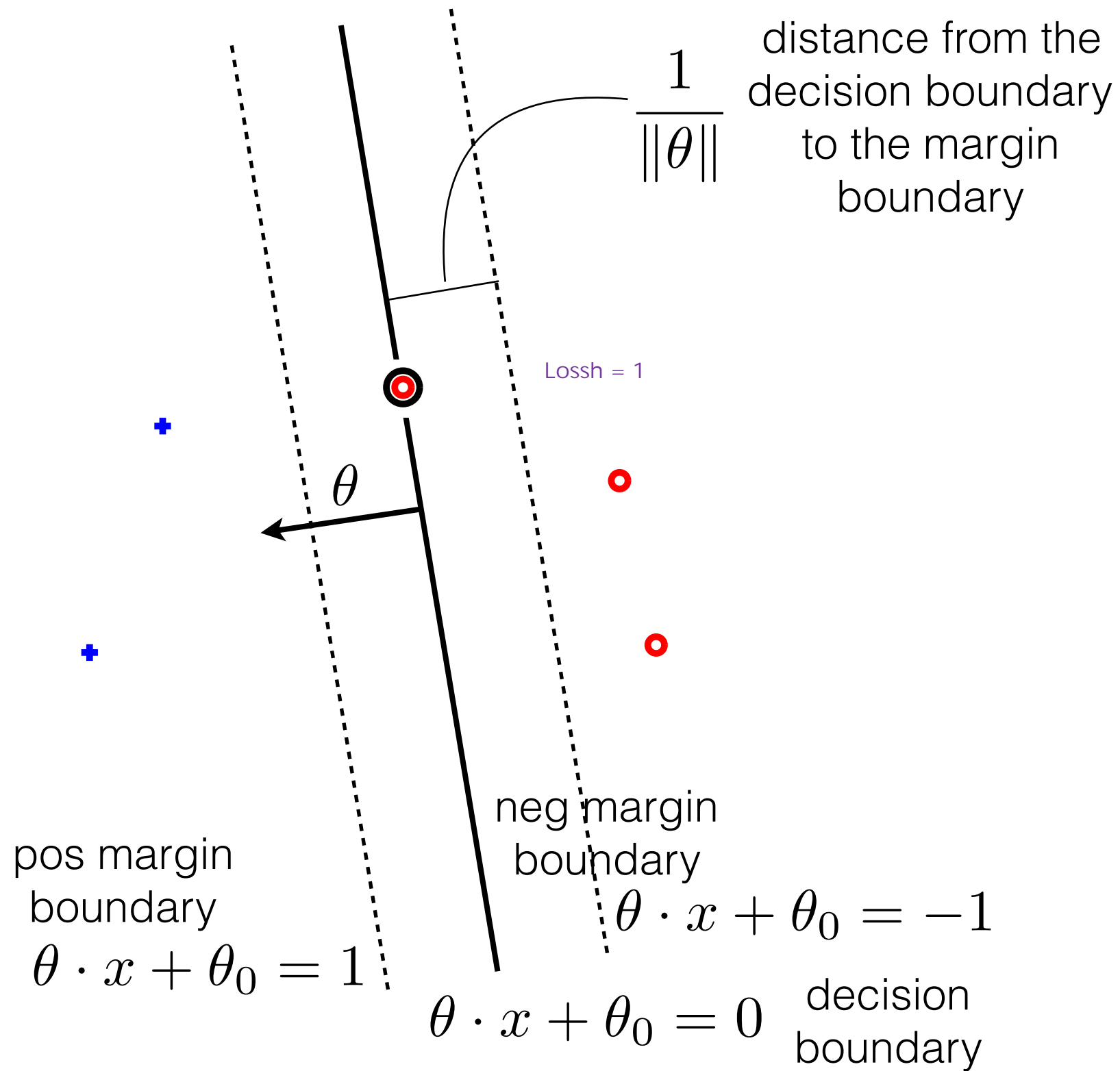$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h\big(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\big) + \frac{\lambda}{2}\|\theta\|^2$$

$$\frac{1}{\|\theta\|}$$

distance from the decision boundary to the margin boundary

$$\theta$$

Point lies in decision boundary
(0 < θ*x + θ_0 < 1)
Lossh(0*0) = 1 - 0 = 1(z < 1)

pos margin boundary
$$\theta \cdot x + \theta_0 = 1$$

neg margin boundary
$$\theta \cdot x + \theta_0 = -1$$

$$\theta \cdot x + \theta_0 = 0$$

decision boundary

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

$$\frac{1}{\|\theta\|}$$

distance from the decision boundary to the margin boundary

Lossh = 1

θ

neg margin boundary

pos margin boundary

$\theta \cdot x + \theta_0 = 1$

$\theta \cdot x + \theta_0 = -1$

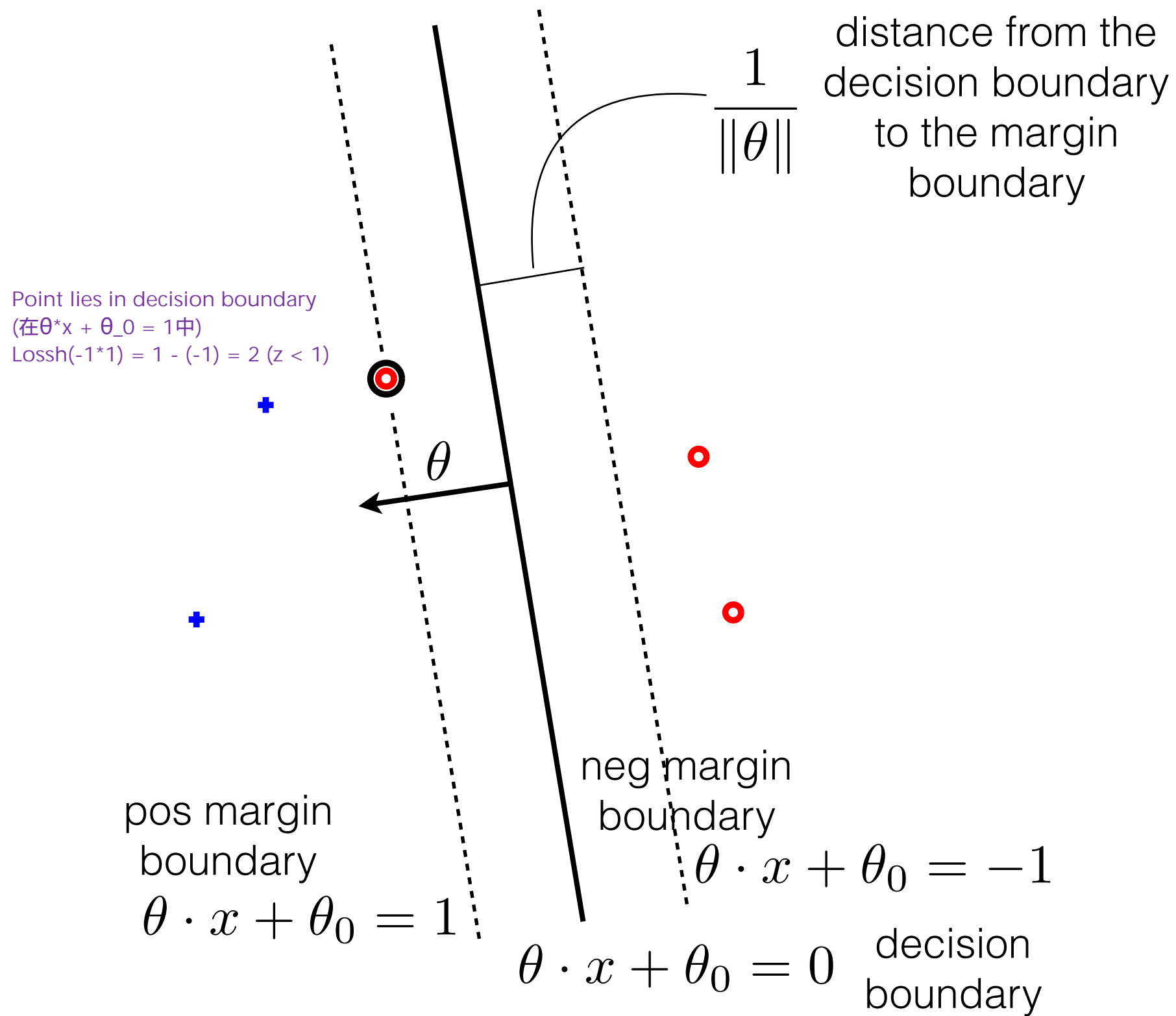$\theta \cdot x + \theta_0 = 0$

decision boundary

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

$$\frac{1}{\|\theta\|}$$

distance from the decision boundary to the margin boundary

Point lies in decision boundary
(θ*x + θ_0 = 1)
Loss(1-1*1) = 1 - (-1) = 2 (z < 1)

θ

neg margin boundary

pos margin boundary

$\theta \cdot x + \theta_0 = 1$

$\theta \cdot x + \theta_0 = 1$

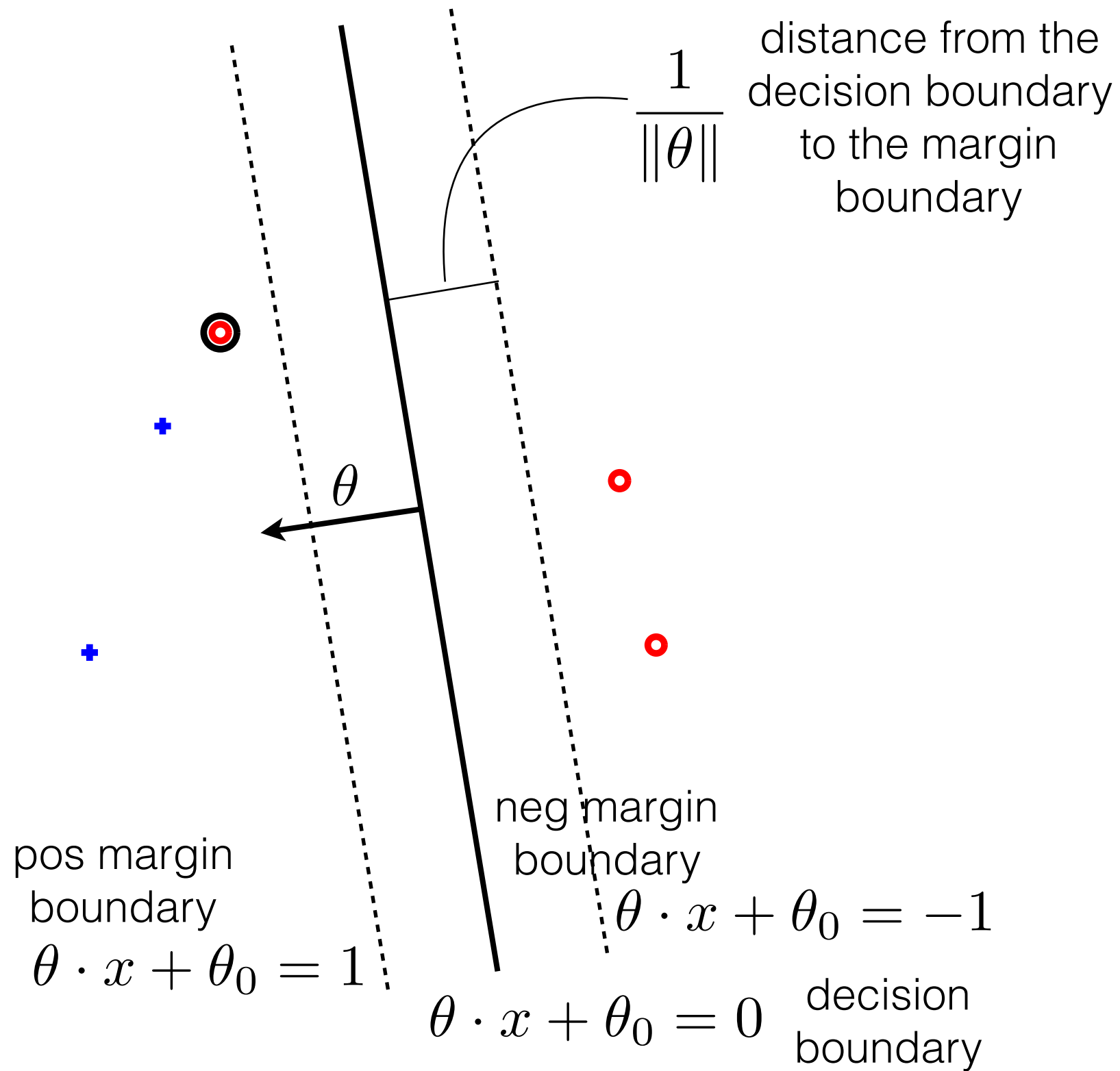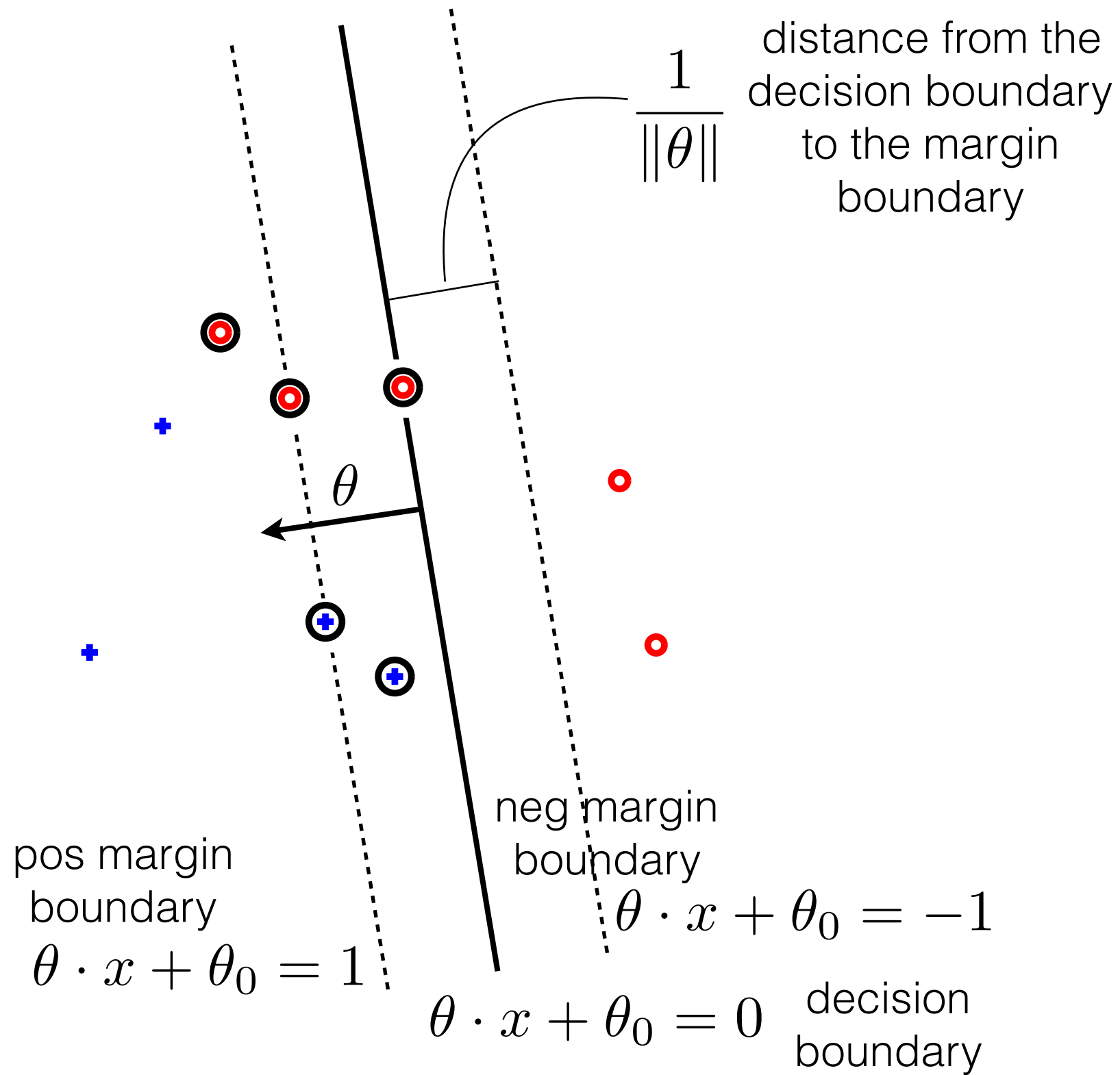$\theta \cdot x + \theta_0 = 0$

decision boundary

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

$\dfrac{1}{\|\theta\|}$ distance from the decision boundary to the margin boundary

θ

neg margin boundary

pos margin boundary

$\theta \cdot x + \theta_0 = -1$

$\theta \cdot x + \theta_0 = 1$

$\theta \cdot x + \theta_0 = 0$

decision boundary

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

$$\frac{1}{\|\theta\|}$$

distance from the
decision boundary
to the margin
boundary

θ

neg margin
boundary

pos margin
boundary

$\theta \cdot x + \theta_0 = 1$

$\theta \cdot x + \theta_0 = -1$

$\theta \cdot x + \theta_0 = 0$

decision
boundary

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

Things to know

- General optimization formulation of learning

  objective function = average loss + regularization

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

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