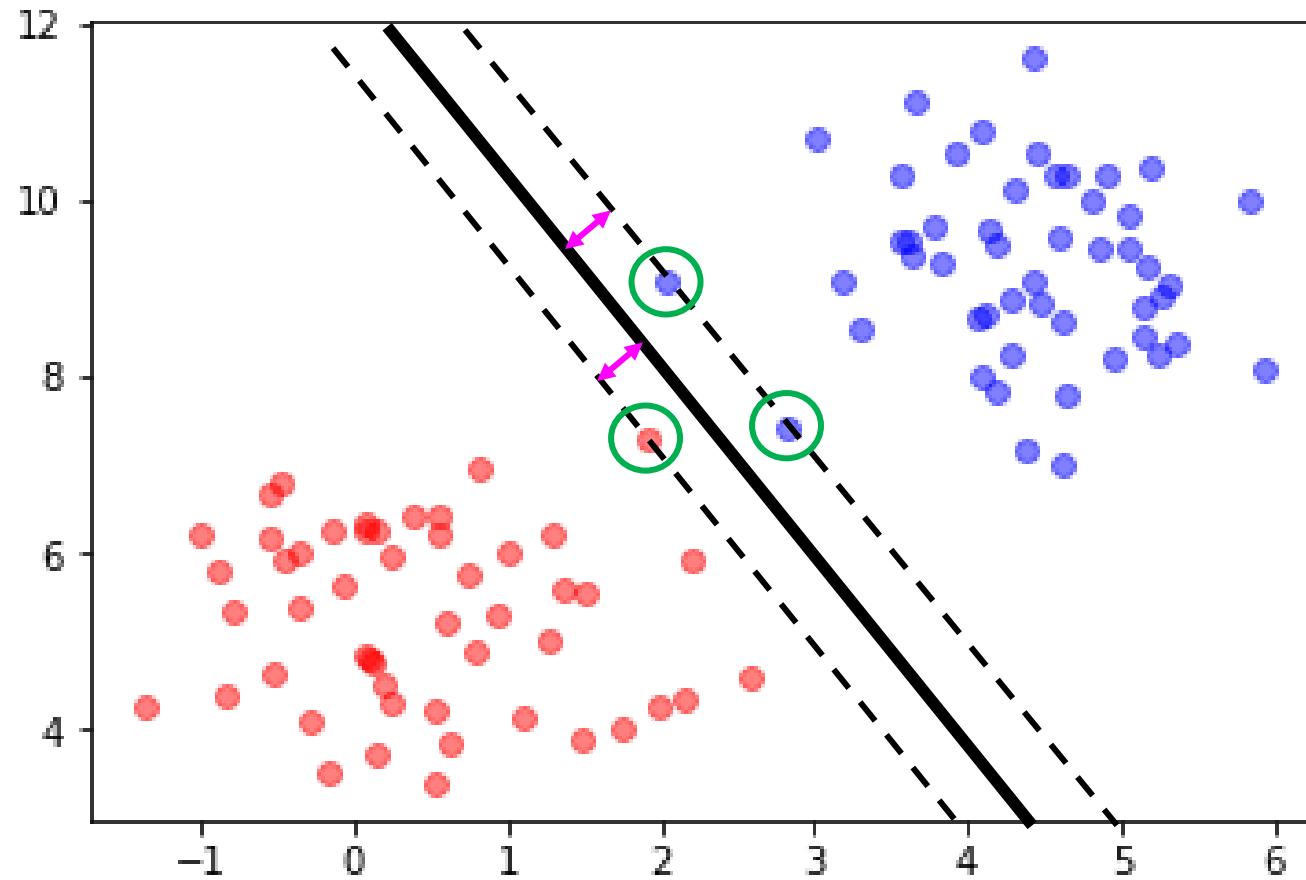


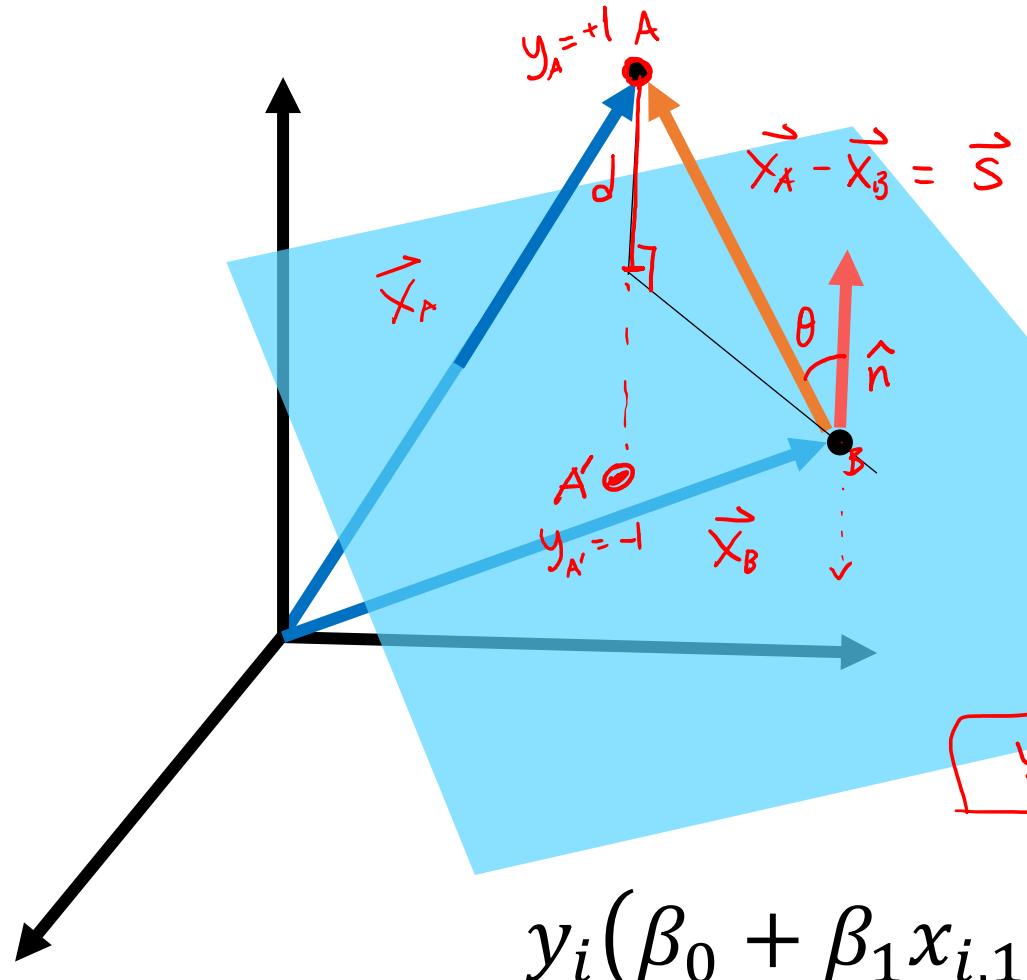
Support Vector Machine

Maximum margin classifier



Support
Margin

Maximum margin classifier



$$d = |s| \cos \theta = \vec{s} \cdot \hat{n}$$

$$\vec{s} = (s_1, s_2, s_3), (s_1, \dots, s_p)$$

$$\hat{n} = (w_1, w_2, w_3)$$

$$(= w_1^2 + w_2^2 + w_3^2)$$

$$d = s_1 w_1 + s_2 w_2 + s_3 w_3$$

$$x_{A1} w_1 + x_{A2} w_2 + x_{A3} w_3$$

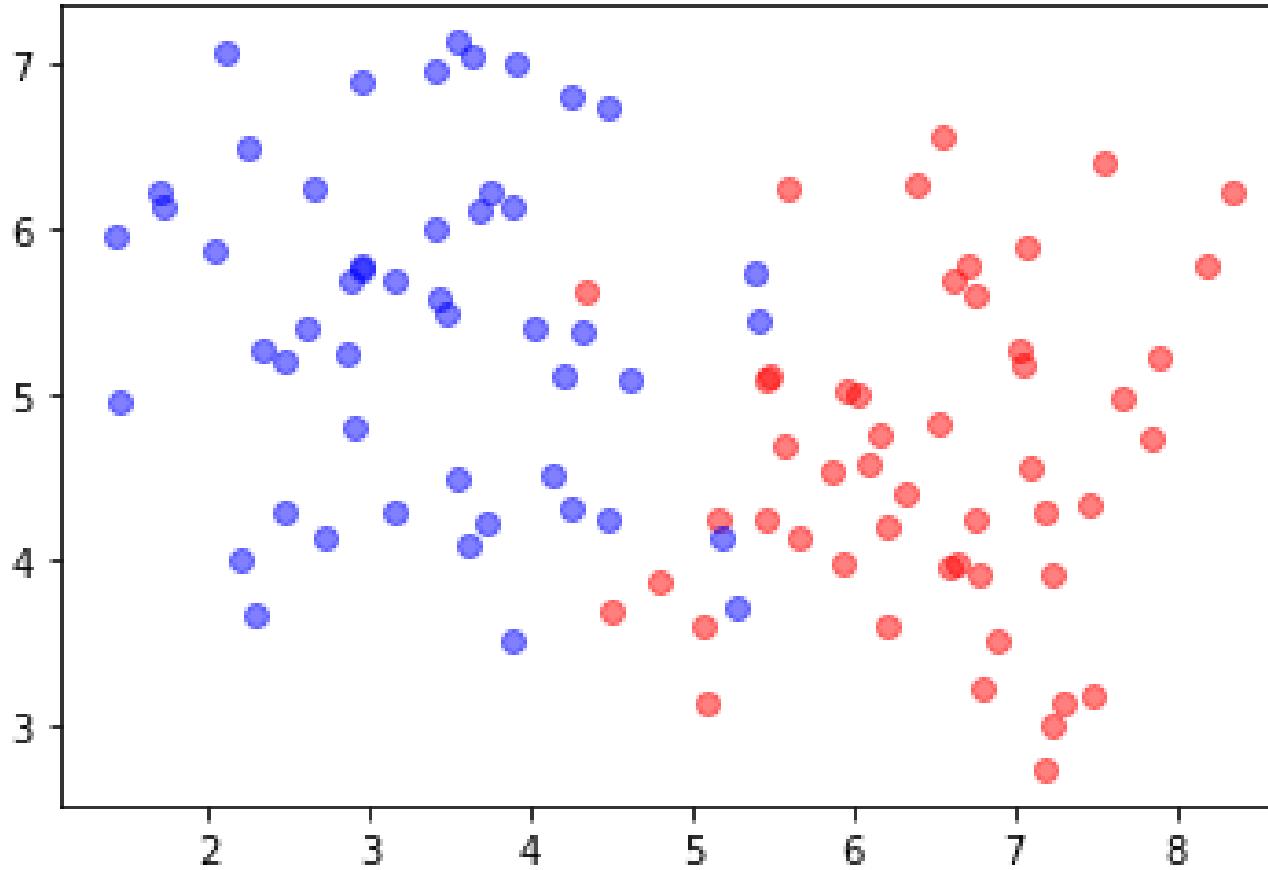
$$- x_{B1} w_1 - x_{B2} w_2 - x_{B3} w_3$$

$$x_A \cdot [x_1 w_1 + x_2 w_2 + x_3 w_3 + b] = d = M$$

$$y_i (x_{i1} w_1 + x_{i2} w_2 + \dots + x_{ip} w_p + b) \geq M$$

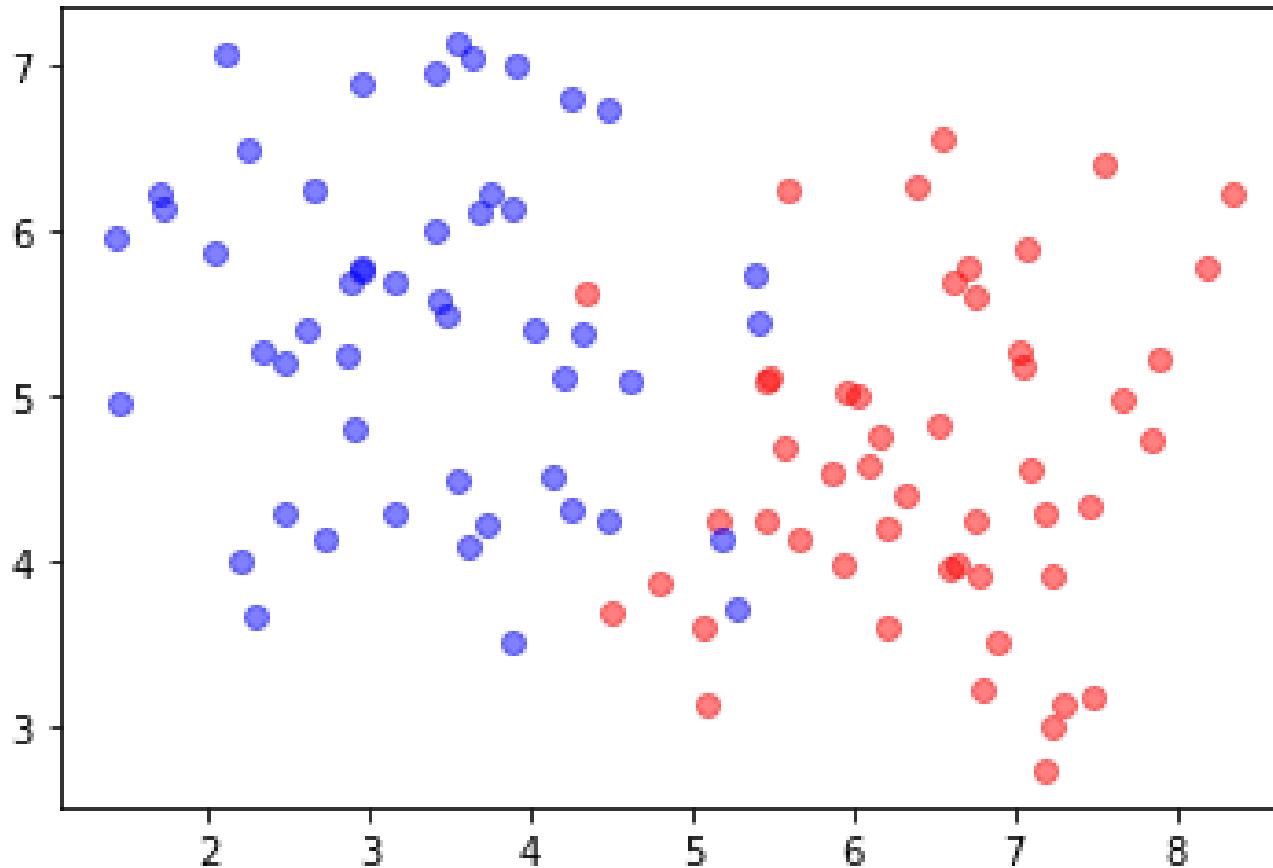
$$y_i (\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \dots) \geq M$$

The impossible case...



Can you separate this with
a hyperplane?

How to deal with an inseparable case



We'll have to accept some errors by softening the margin

“soft margin classifier”

or called

“support vector classifier”

Soft margin classifier

$$y_i(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \dots) \geq M(1 - \epsilon_i)$$

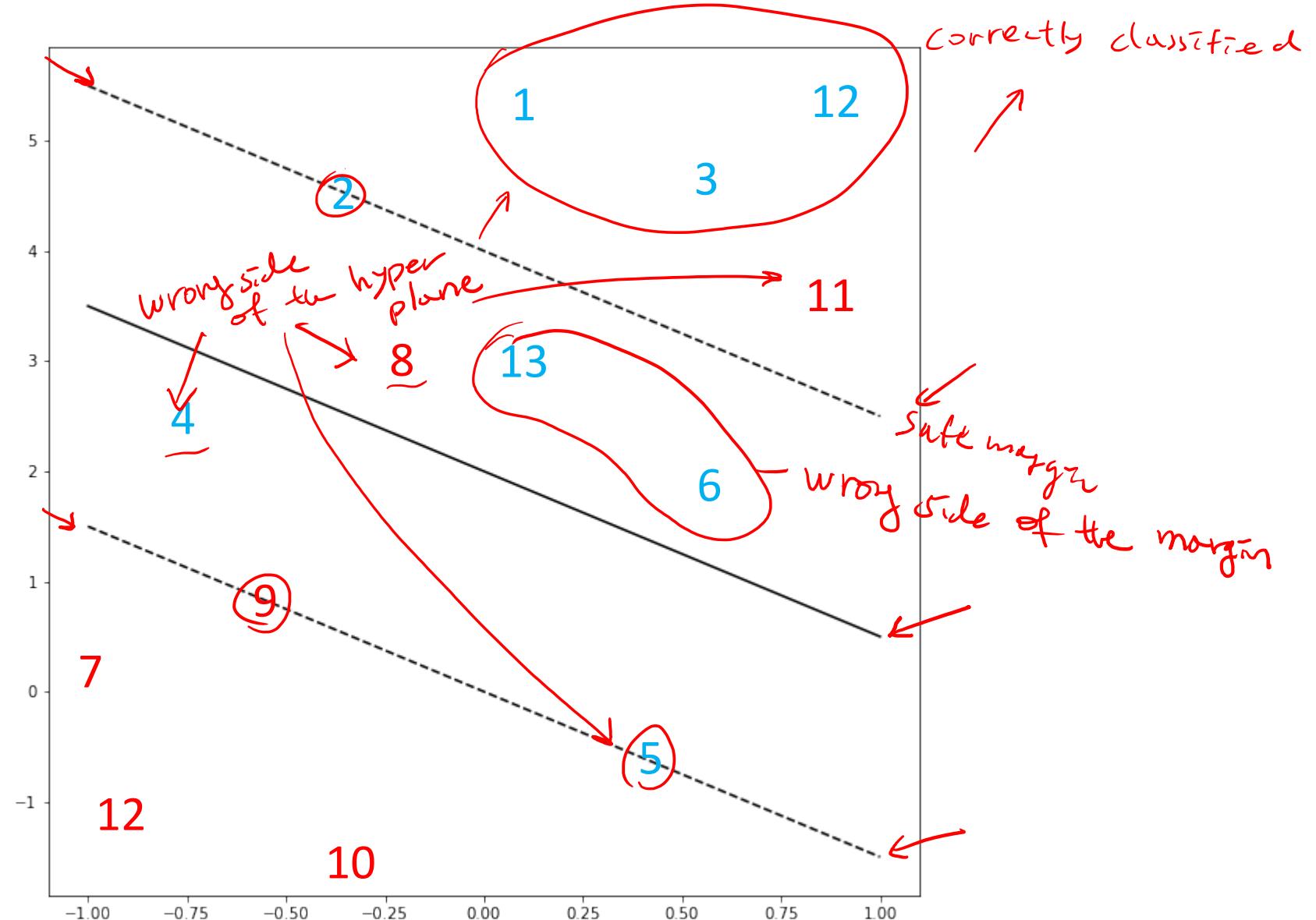
$\stackrel{= x_1}{-1}$

$$\sum_j^p \beta_j^2 = 1$$

$$\underline{\epsilon_i \geq 0}$$

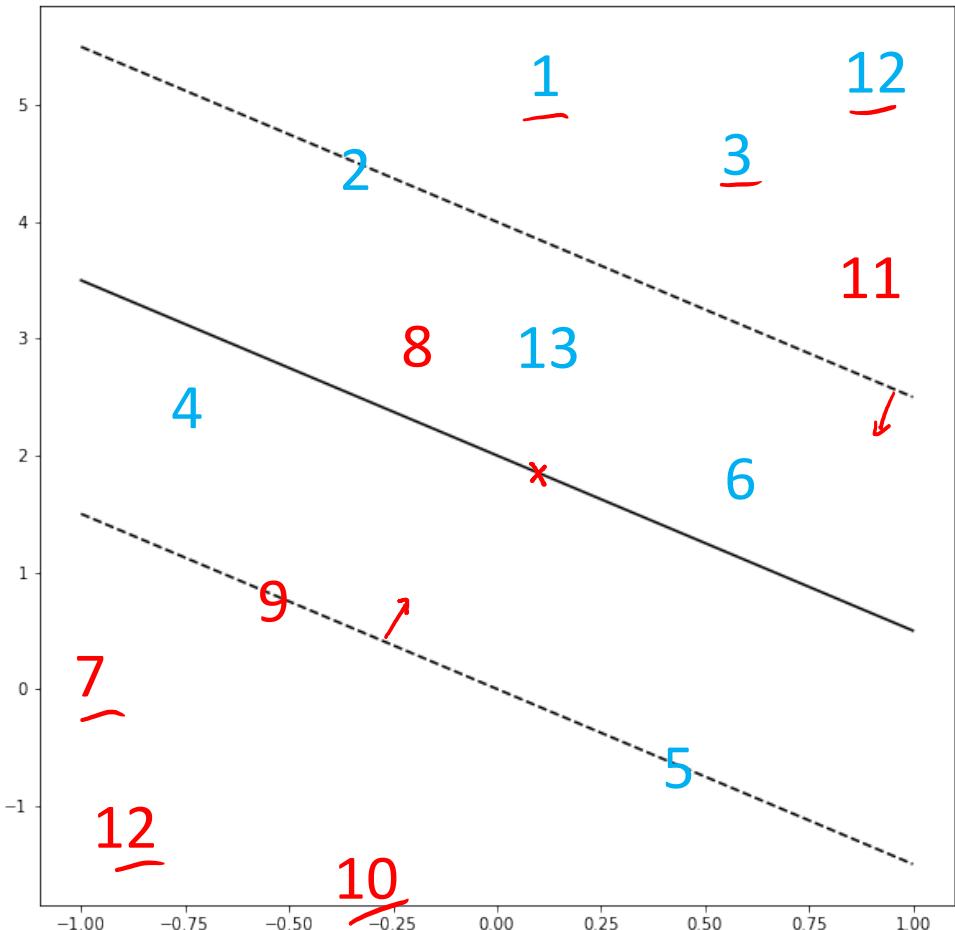
$$\sum_i^n \epsilon_i \leq \overset{\text{Budget}}{\circled{C}}$$

Soft margin classifier



Quiz: Soft margin classifier

$$\underline{y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i)}$$
$$\underline{\epsilon_i \geq 0}$$



- Correct side of the margin

$$\xi_i = 0$$

- wrong side of the margin

$$0 < \xi_i \leq 1$$

- wrong side of the hyperplane

$$1 < \xi_i$$

The role of C parameter

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i)$$

$$\epsilon_i \geq 0$$

$$\sum_{i=1}^n \epsilon_i \leq C$$

C is an error budget

C bounds both number and severity of violations

C is a hyperparameter

The role of C parameter

- Q1. What's the maximum number of supports on the wrong side the hyperplane given C?
- Q2. What happens to the margin M when C decreases?
- Q3. What happens to the bias and variance when C is small?

The role of C parameter

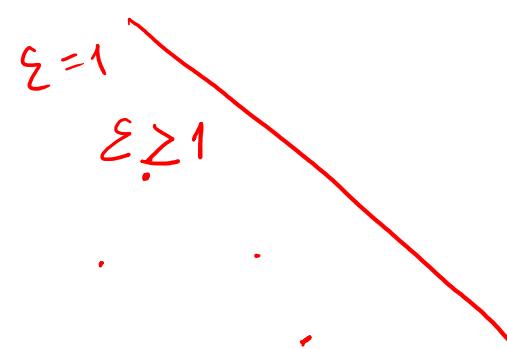
Q1. What's the maximum number of supports on the wrong side the hyperplane given C?

ANS: C

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \underbrace{\epsilon_i}_{\geq 0})$$

$$\epsilon_i \geq 0$$

$$\sum_{i=1}^n \epsilon_i \leq \underline{C}$$



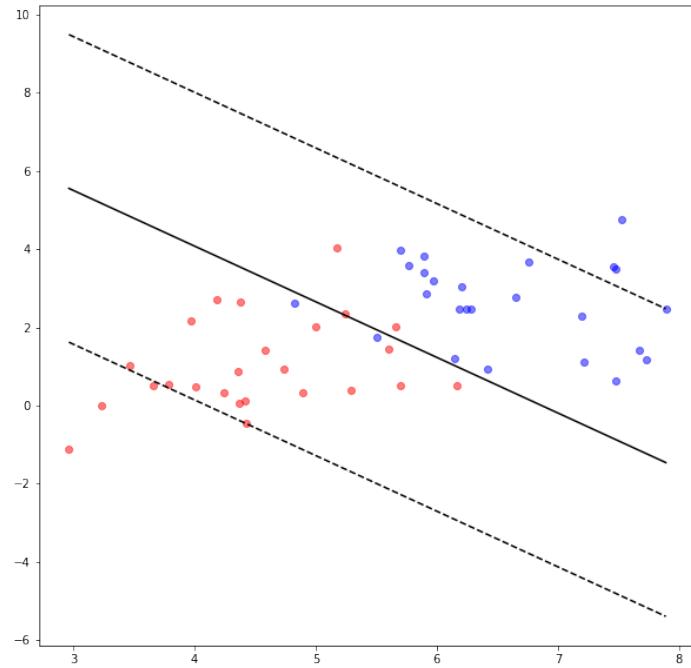
The role of C parameter

Q2. What happens to the margin when C decreases?

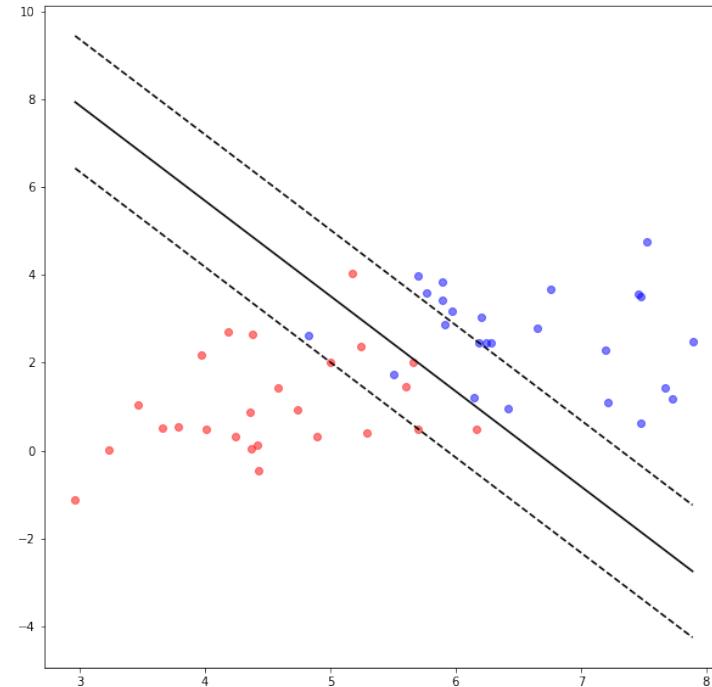
$$\sum_{i=1}^n \epsilon_i \leq C$$

C

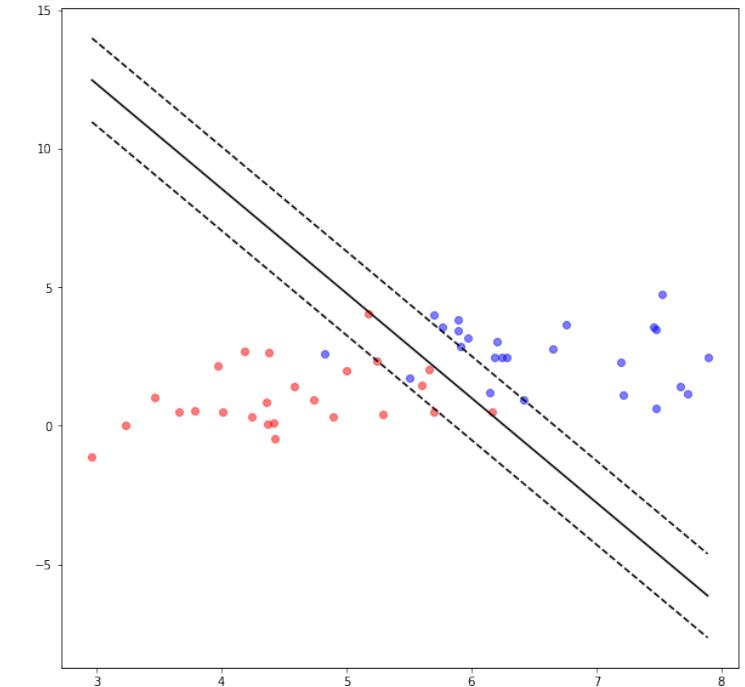
A



B



✓



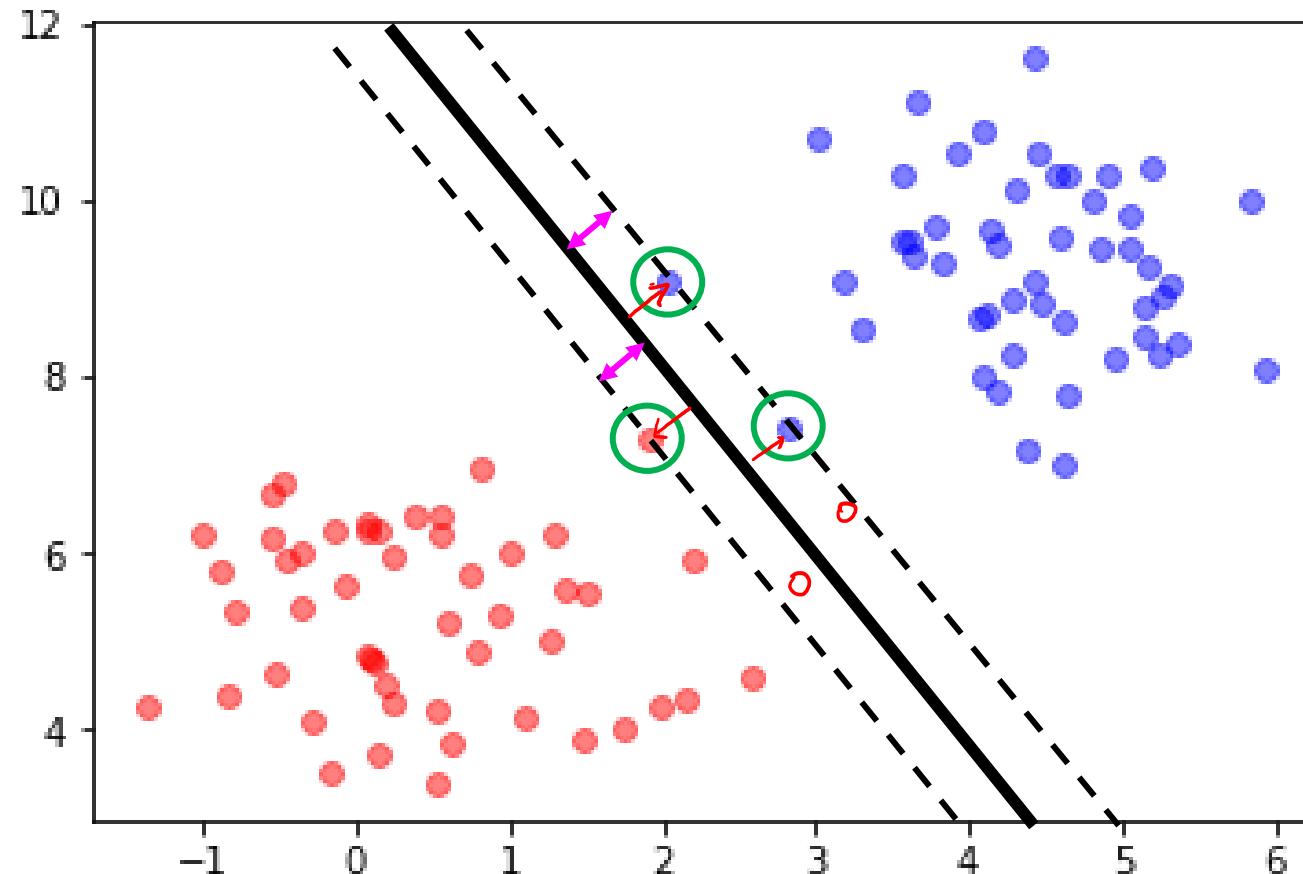
ANS: the margin becomes narrower

The role of C parameter

Q3. What happens to the bias and variance when C is small?

ANS: small C gives lower bias and higher variance

Recap



Support

Margin

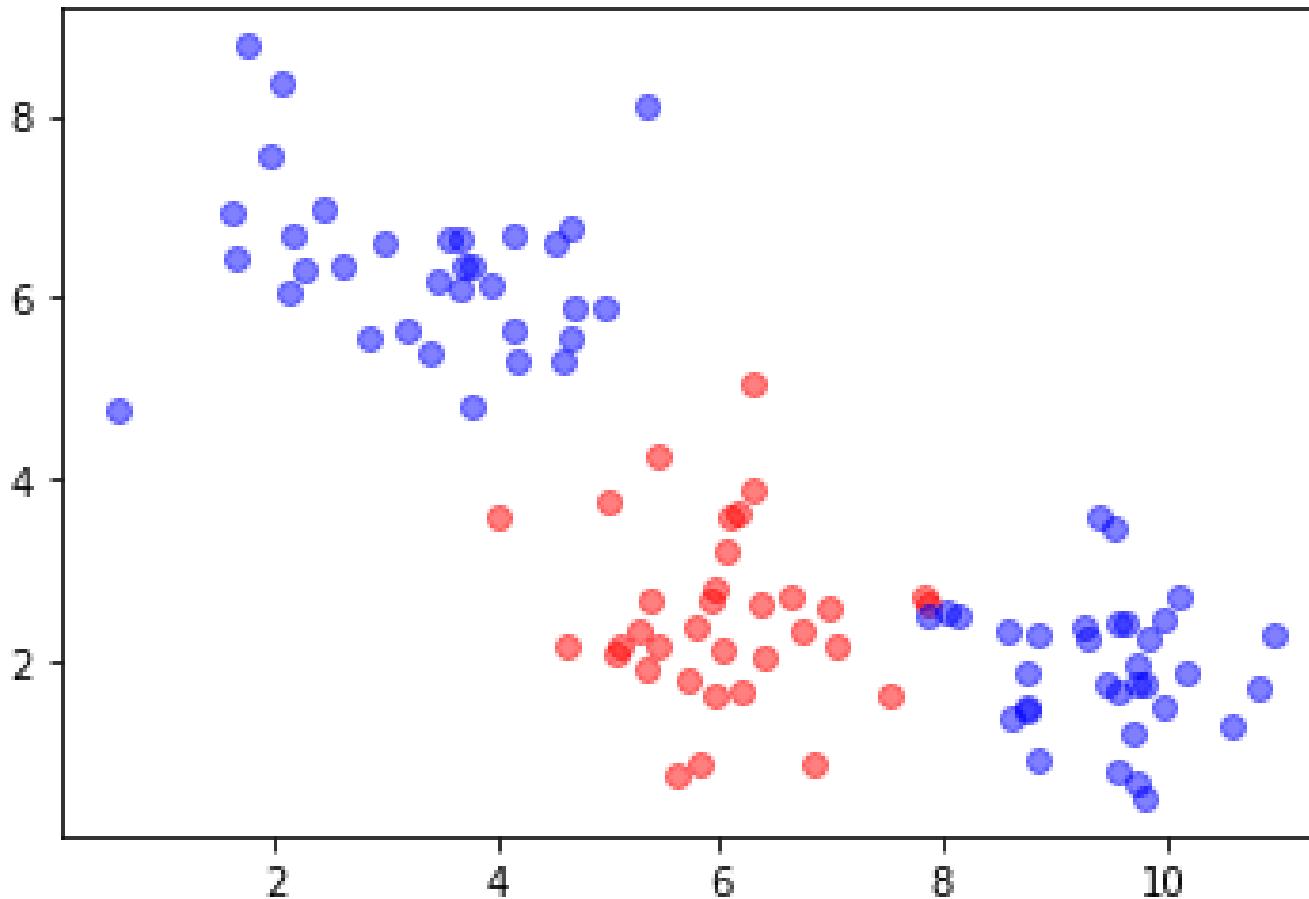
$$y_i(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \dots) \geq M(1 - \epsilon_i)$$

$$\epsilon_i \geq 0$$

$$\sum_{i=1}^n \epsilon_i \leq C$$

Beyond linearly separable data

How can we separate this kind of data with SVC?



$$X^T W + b$$