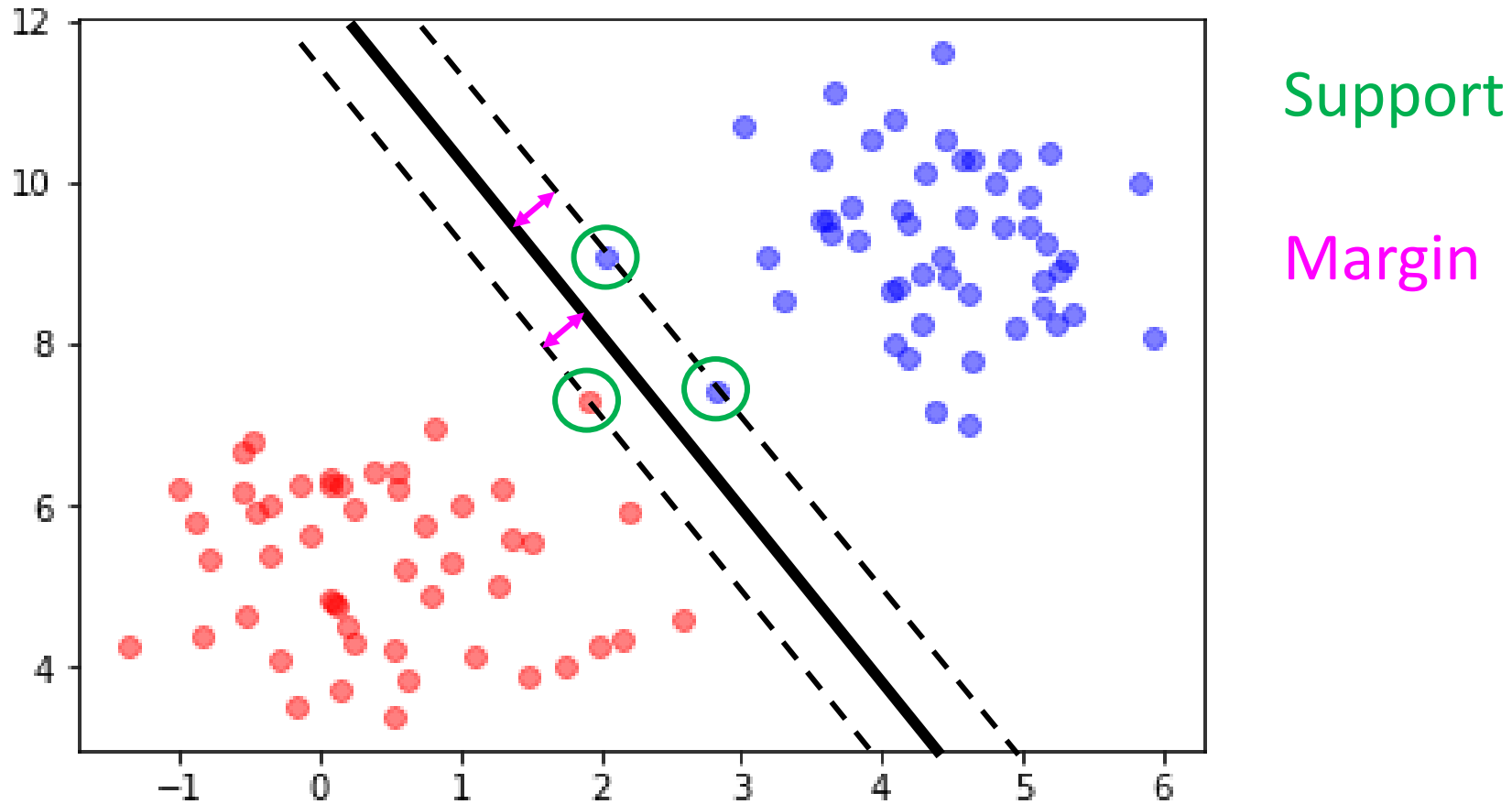


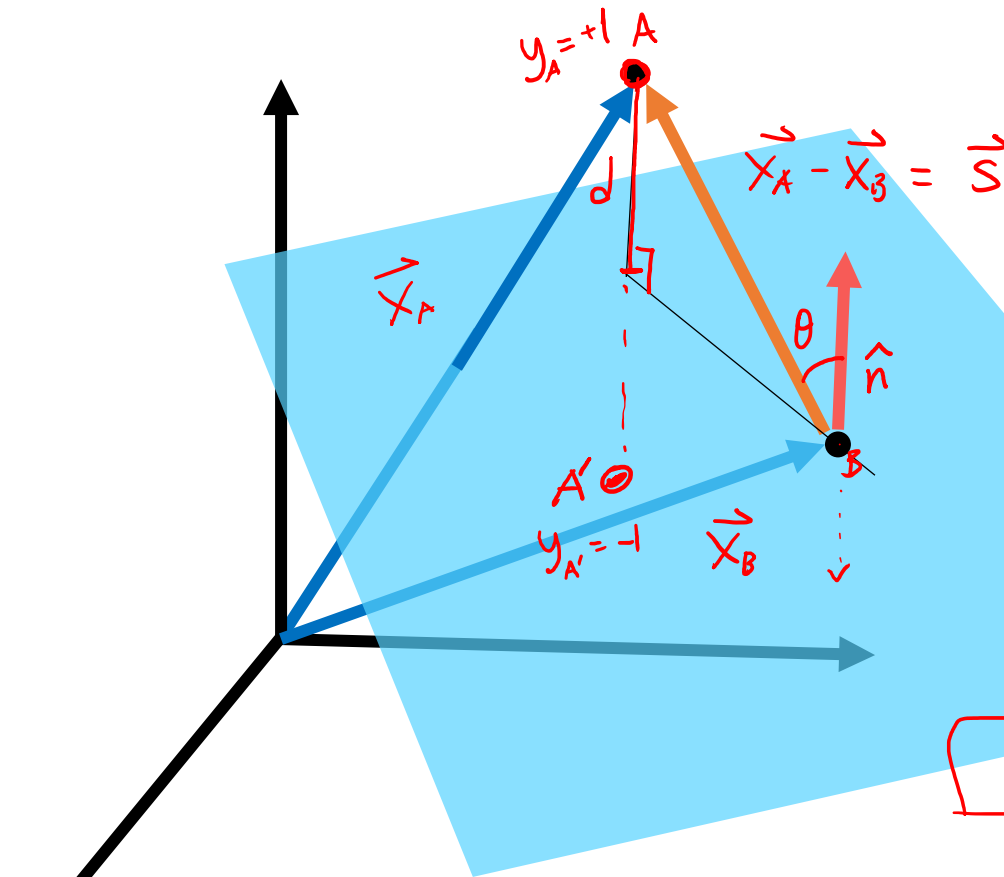
Support Vector Machine



Maximum margin classifier



Maximum margin classifier



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

$$\vec{S} = (S_1, S_2, S_3), \quad (S_1, \dots, S_p)$$

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

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

$$d = S_1 w_1 + S_2 w_2 + S_3 w_3$$

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

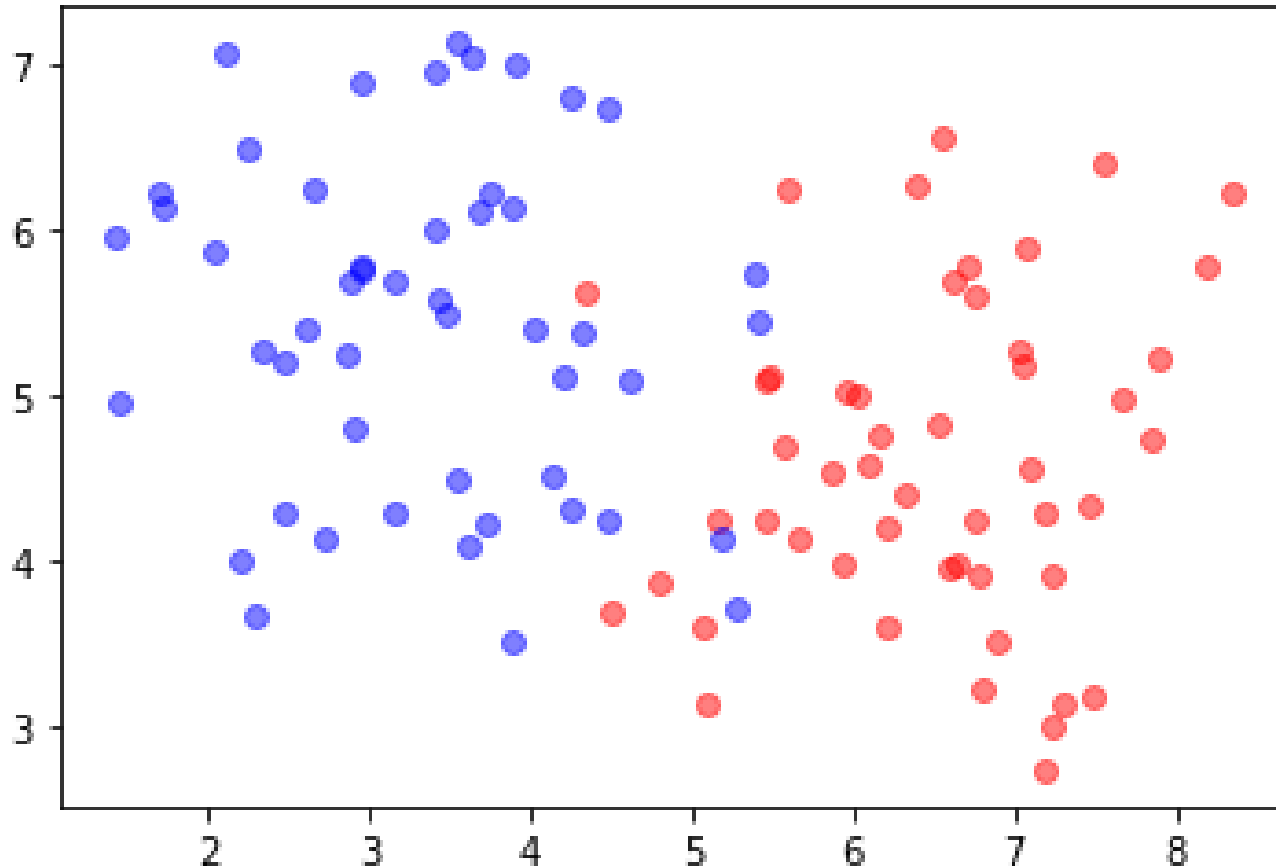
$$- X_{B1} w_1 - X_{B2} w_2 - X_{B3} w_3$$

$$X_A = X \quad \boxed{X_1 w_1 + X_2 w_2 + X_3 w_3 + b} = d = M$$

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

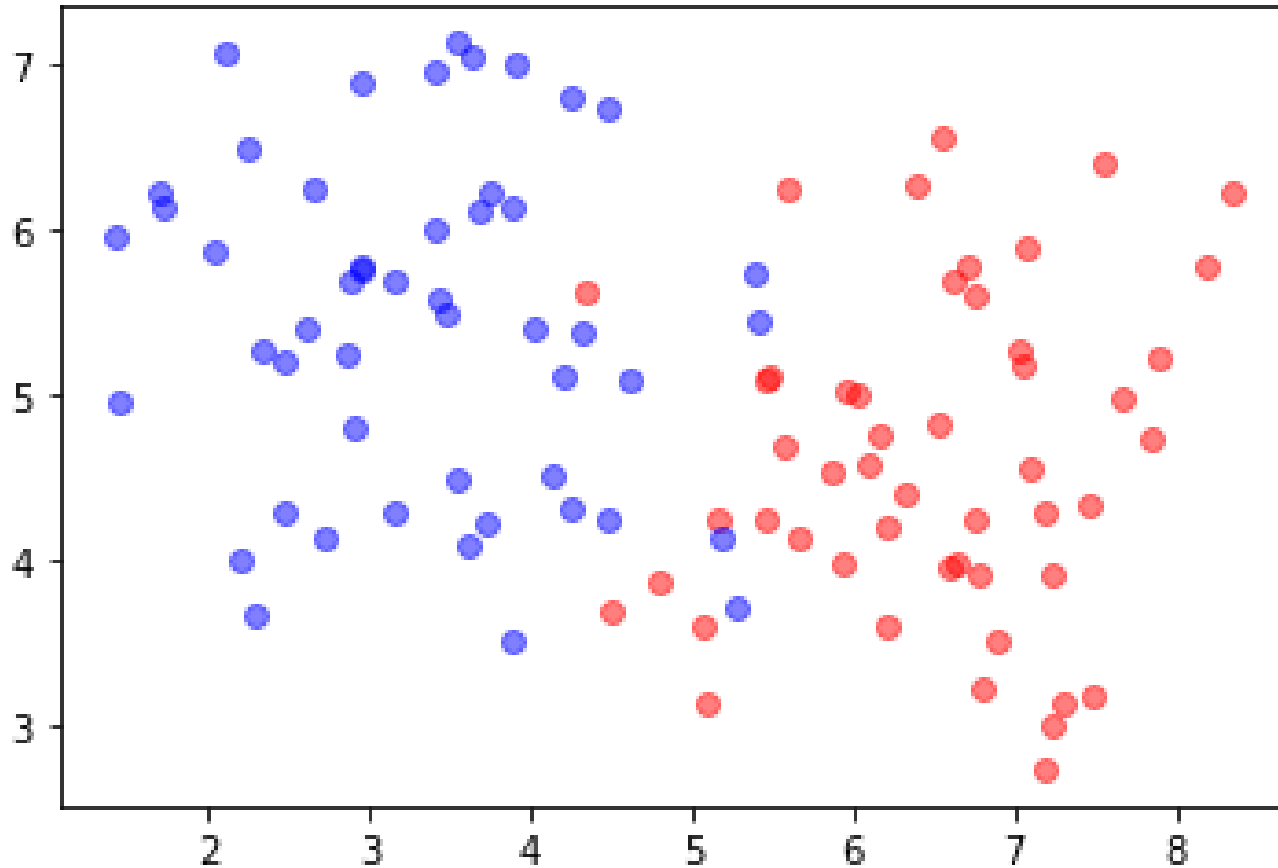
$$y_i (\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \dots) \geq M$$

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”

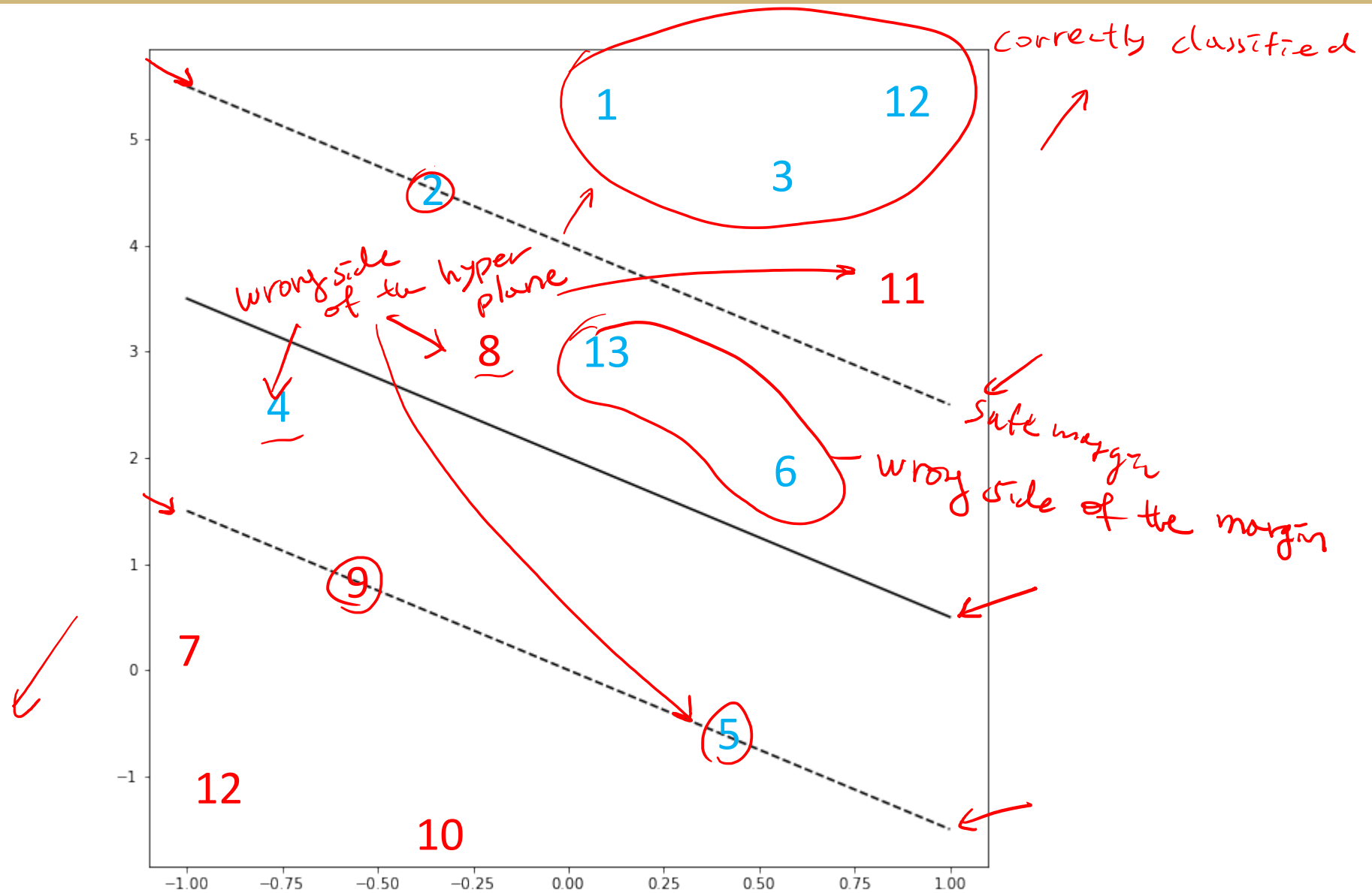
Soft margin classifier

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

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

$$\underbrace{\epsilon_i \geq 0} \quad \underbrace{\sum_i^n \epsilon_i \leq C}_{\text{Budget}}$$

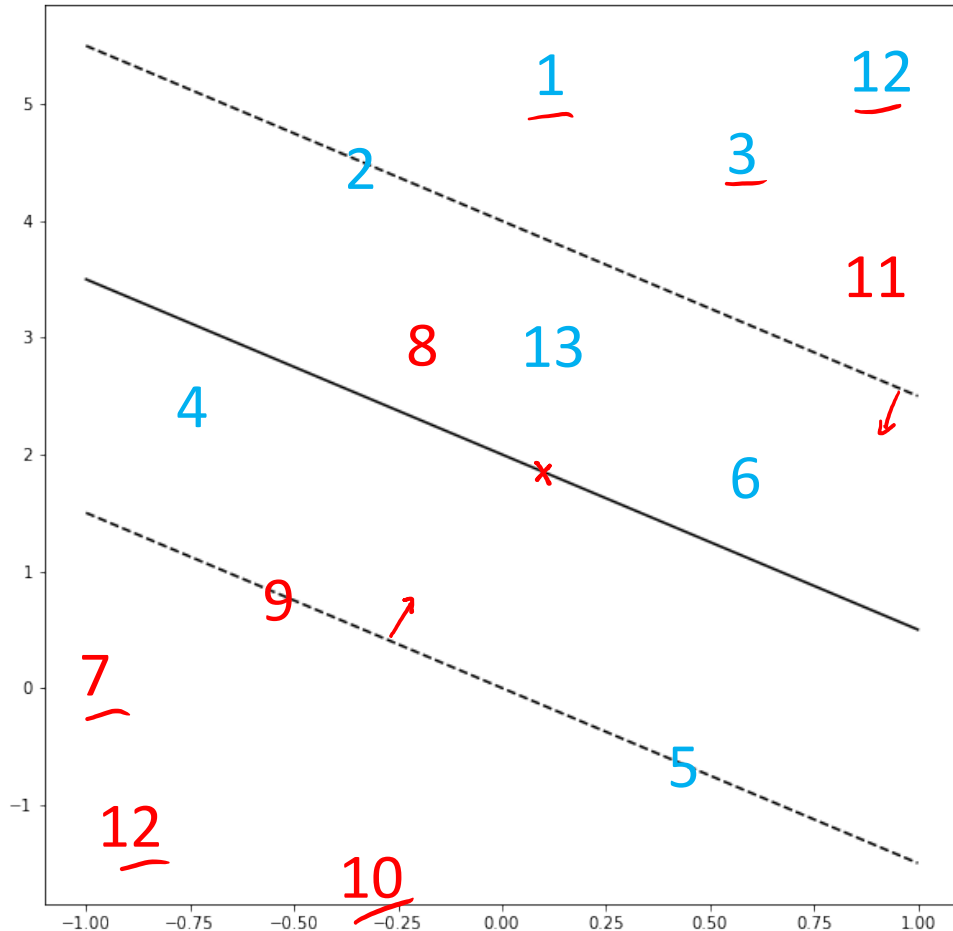
Soft margin classifier



Quiz: Soft margin classifier

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

$$\epsilon_i \geq 0$$



- Correct side of the margin

$$\epsilon_i = 0$$

- wrong side of the margin

$$0 < \epsilon_i \leq 1$$

- wrong side of the hyperplane

$$1 < \epsilon_i$$

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

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

$$\epsilon_i \geq 0$$

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

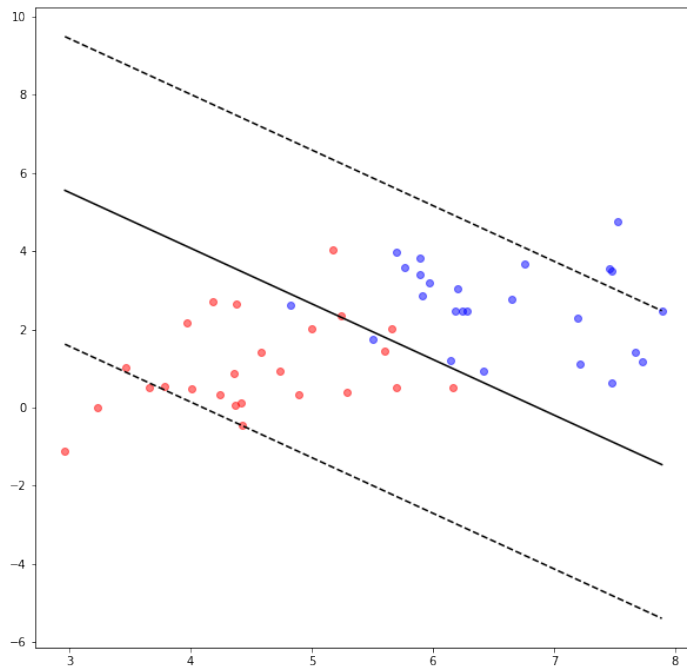
$\xi = 1$
 $\xi \geq 1$
.

The role of C parameter

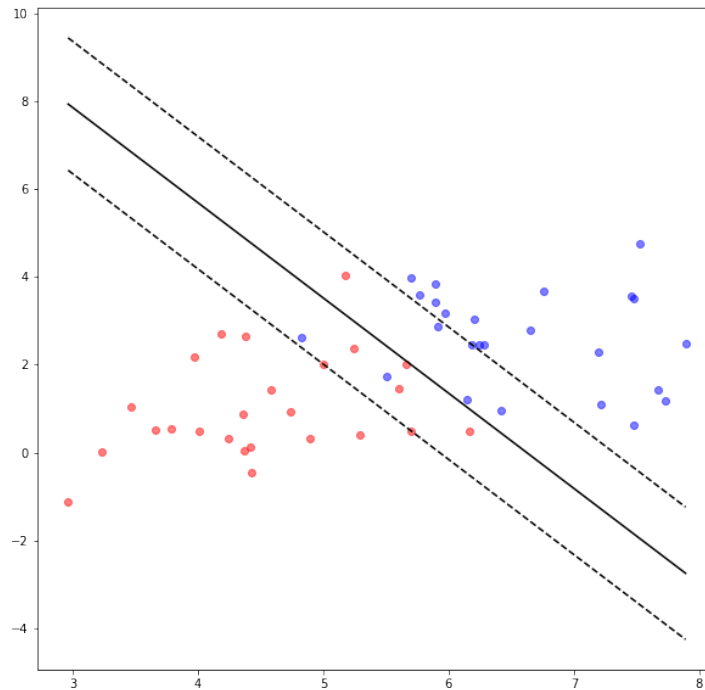
Q2. What happens to the margin when C decreases?

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

A

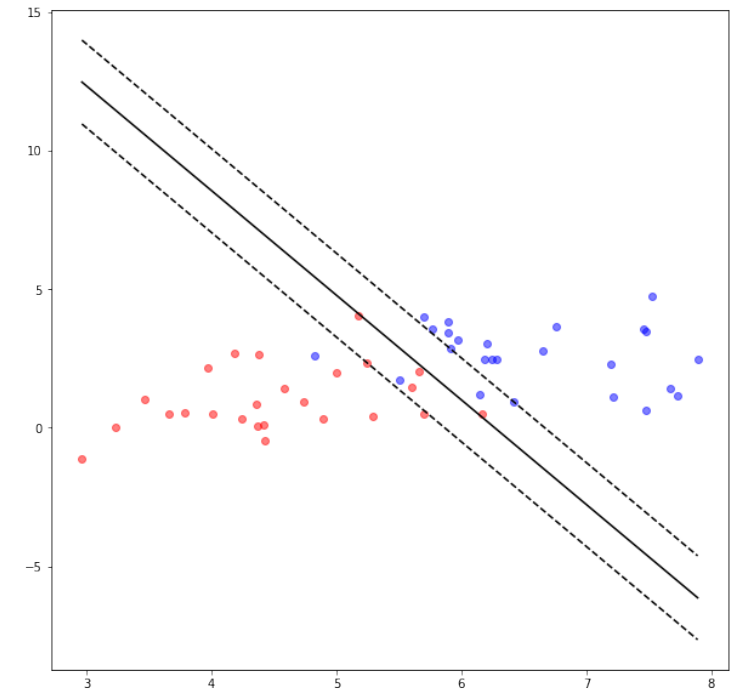


B



✓

C



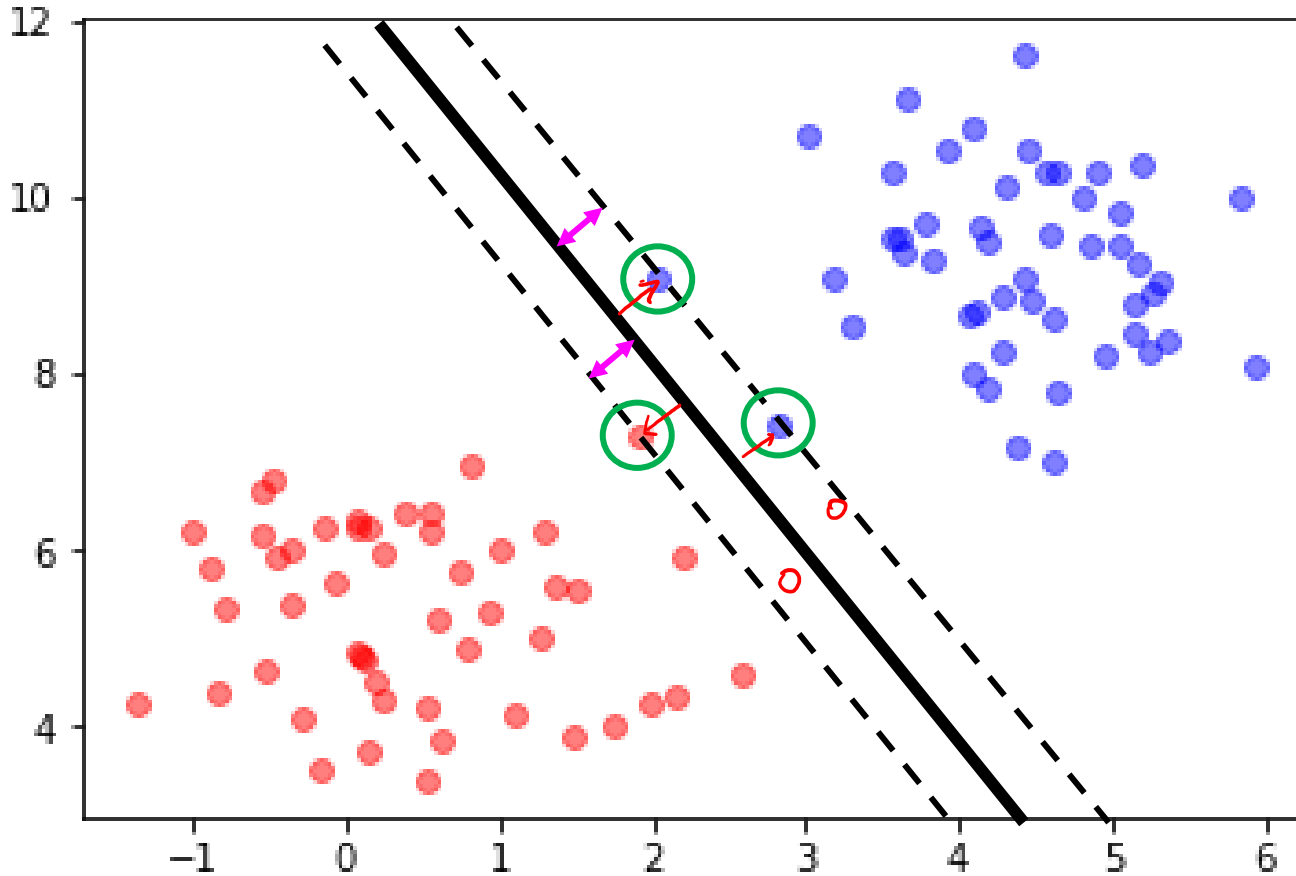
ANS: the margin becomes narrower

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



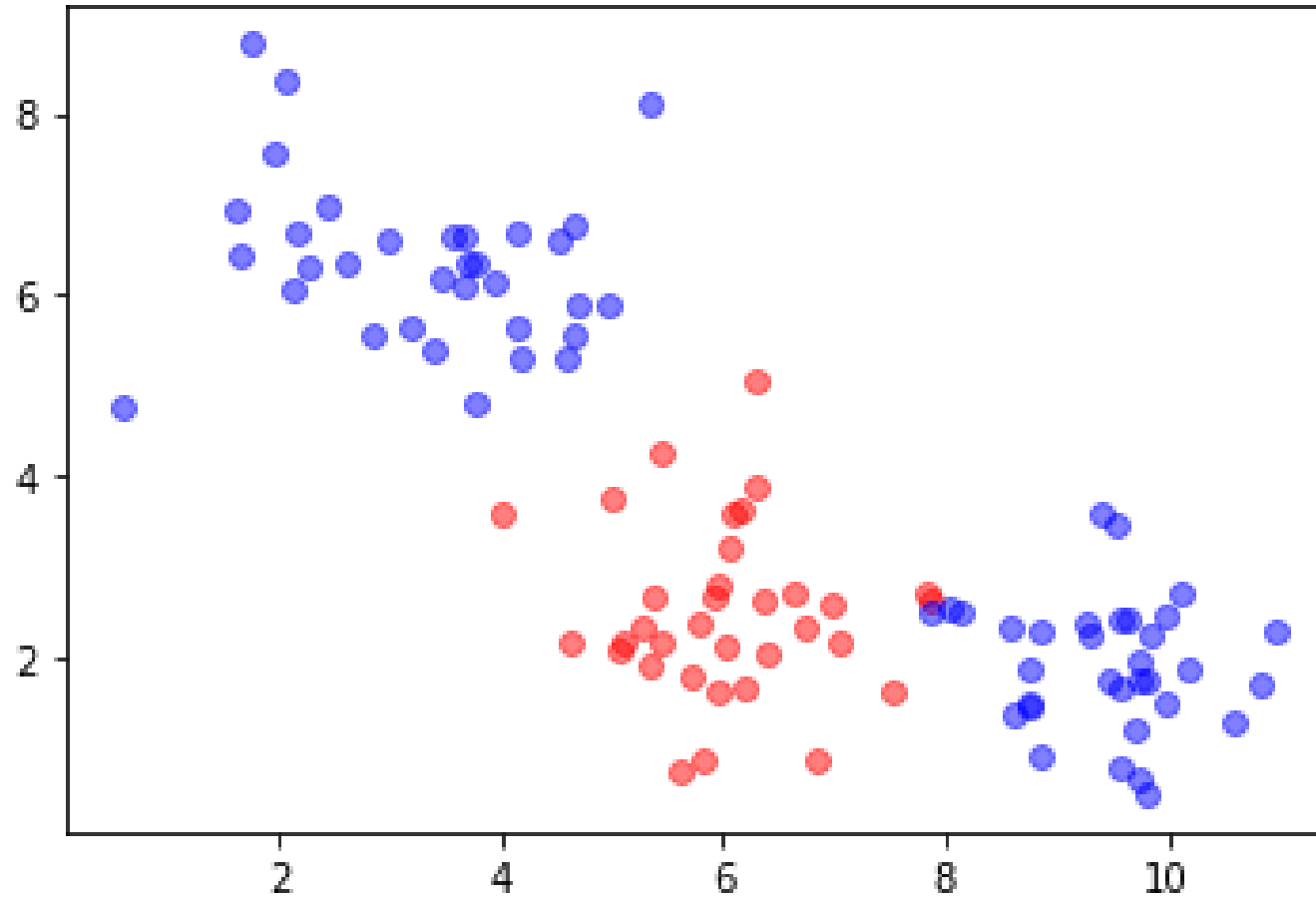
Support

Margin

$$y_i(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \dots) \geq M(1 - \epsilon_i) \quad \epsilon_i \geq 0 \quad \underline{\sum_{i=1}^n \epsilon_i \leq C}$$

Beyond linearly separable data

How can we separate this kind of data with SVC?



$$\underline{X^T W + b}$$