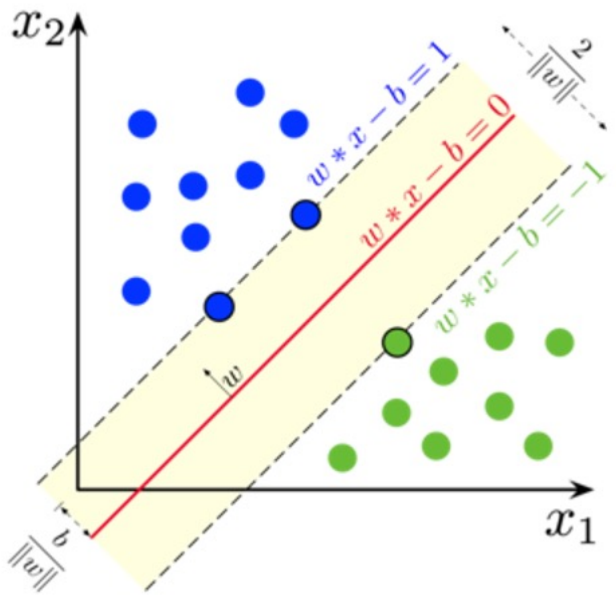
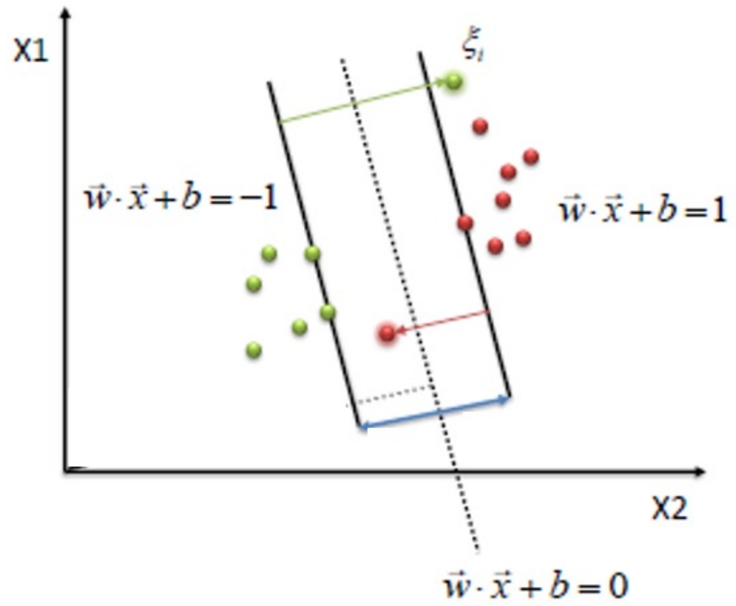
Part 24- SVM classification (Hard and Soft margin)

Prof. Pedram Jahangiry





Topics

Part 23

- SVM Geometry
- SVM Motivation

Part 24

- Maximum Margin Classifier (MMC)
- Support Vector Classifiers (SVC)

Part 25

- Support Vector Machines (SVM)

Part 26

- Support Vector Regressors (SVR)

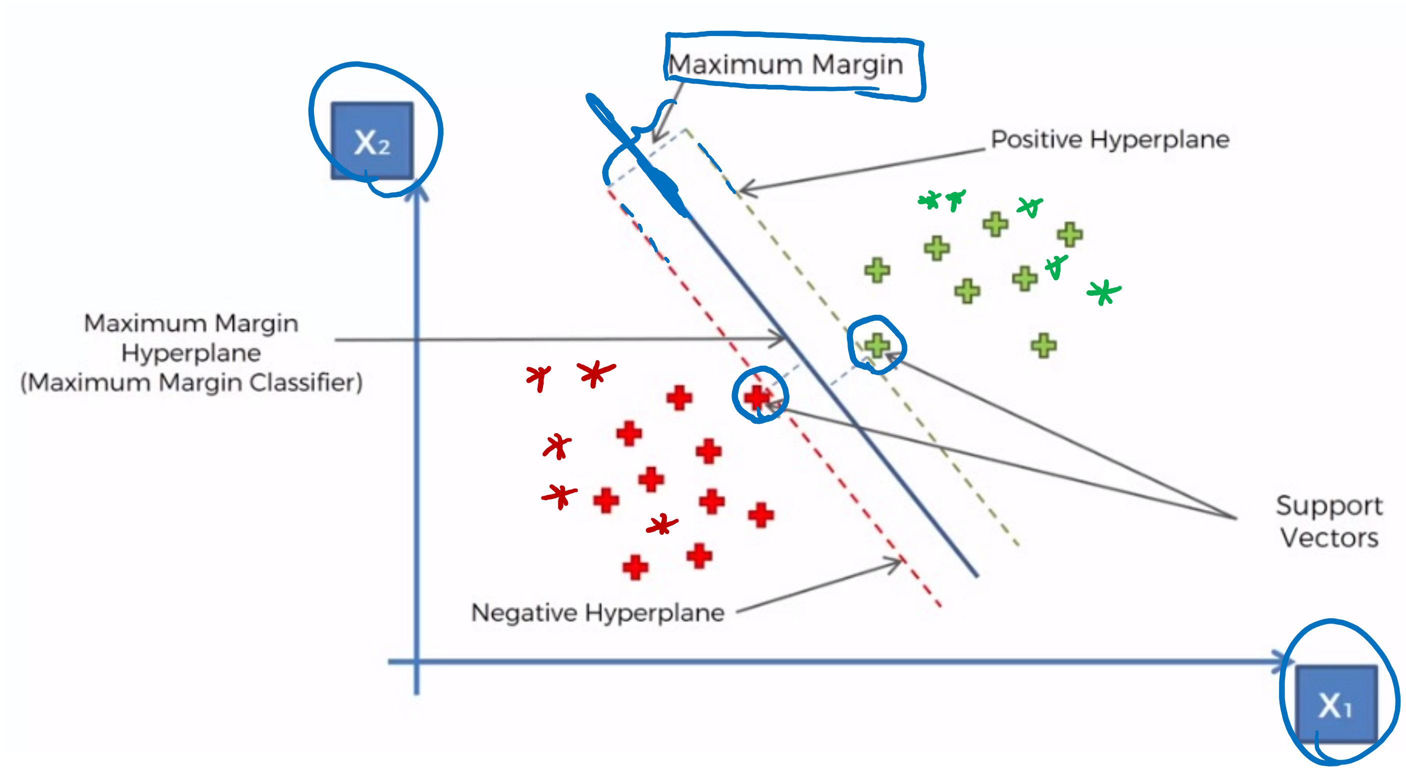
Part 27

- Multiple class classification
- SVM pros and cons
- SVM applications in Finance

Pedram Jahangiry

Maximum Margin Classifier (MMC) – Hard Margin

MMC is the hyperplane that among all separating hyperplanes, find the one that makes the biggest gap (margin) between two classes.



MMC optimization problem

- The core idea of hard margin is to maximize the margin, under the constraint that the classifier does not make any mistake.
- SVMs try to pick the most robust model (by finding the w^* and b^*) among all those that yield a correct classification. If we numerically define blue circles as +1 and green circles as -1, any **good** linear model is expected to satisfy:

$$\begin{cases} y_i\left(\sum_{k=1}^{K} w_k x_{i,k} + b \geq +1\right) & \text{when } y_i = +1 \\ y_i\left(\sum_{k=1}^{K} w_k x_{i,k} + b \leq -1\right) & \text{when } y_i = -1 \end{cases}$$

$\geq$

$$\underset{w,b}{\text{Min}} \; \frac{1}{2}||w||^2$$

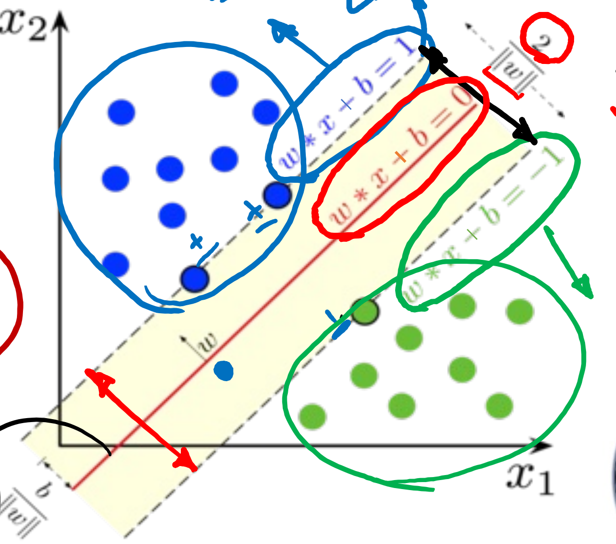
$$s.t. \; y_i\left(\sum_{k=1}^{K} w_k x_{i,k} + b\right) \geq 1$$

$\frac{2}{||w||}$

$|w|$

$\hat{f}(w,b,x)$

blue mg

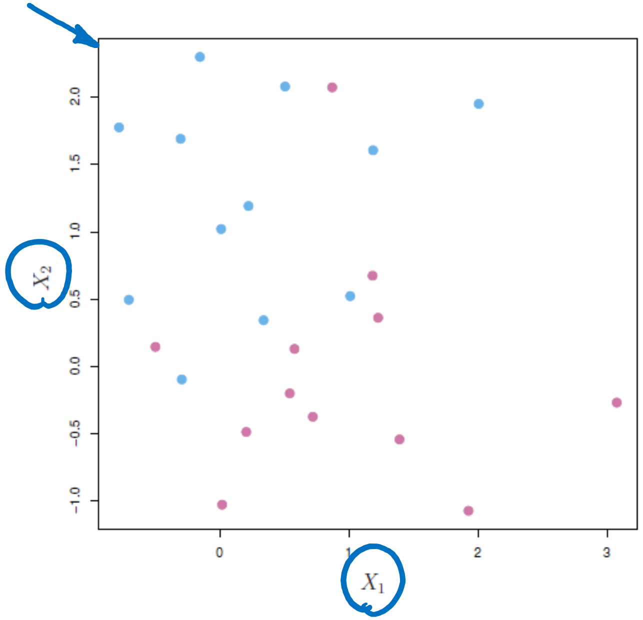


$\frac{|\hat{f}(x)|}{\sqrt{w_1^2 + w_2^2}}$

L_2

Why should we go beyond the Hard margin?

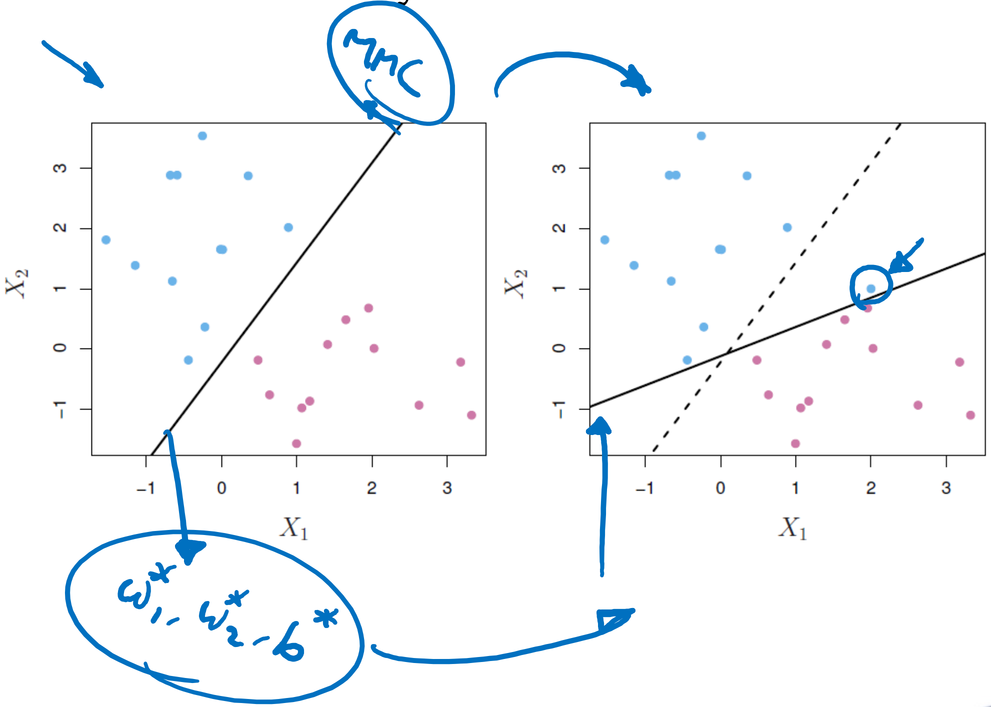
1- The data is non-separable (overlap)



- The MMC optimization problem becomes infeasible whenever the condition cannot be satisfied, that is, when a simple line cannot perfectly separate the labels, no matter the choice of coefficients.

2- The data is noisy

MMC is very sensitive to outliers

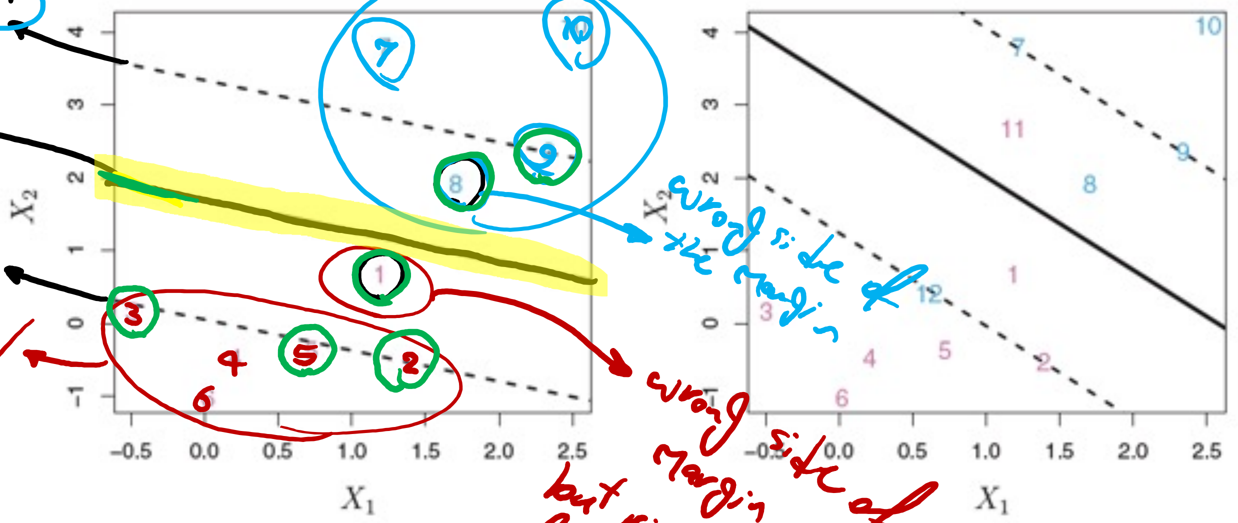


Support Vector Classifier (SVC) – Soft Margin

MMC / hard
SVC / soft

- **Solution**: we can extend the concept of a separating hyperplane in order to develop a hyperplane that **almost** separates the classes, using a so-called **soft margin**.
- The generalization of the maximal margin classifier using soft margin is known as the support vector classifier (SVC).
- It could be worthwhile to misclassify a few training observations in order to do a better job in classifying the remaining observations.

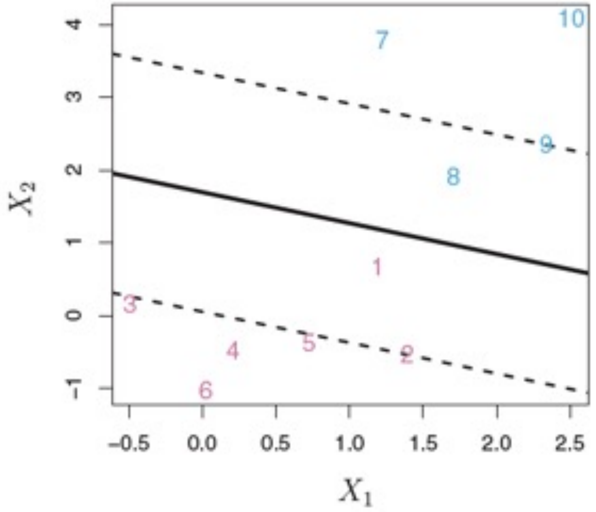
Margin
hyperplane
margin
1, 2, 3, 5, 8, 9
Support vectors
More stable

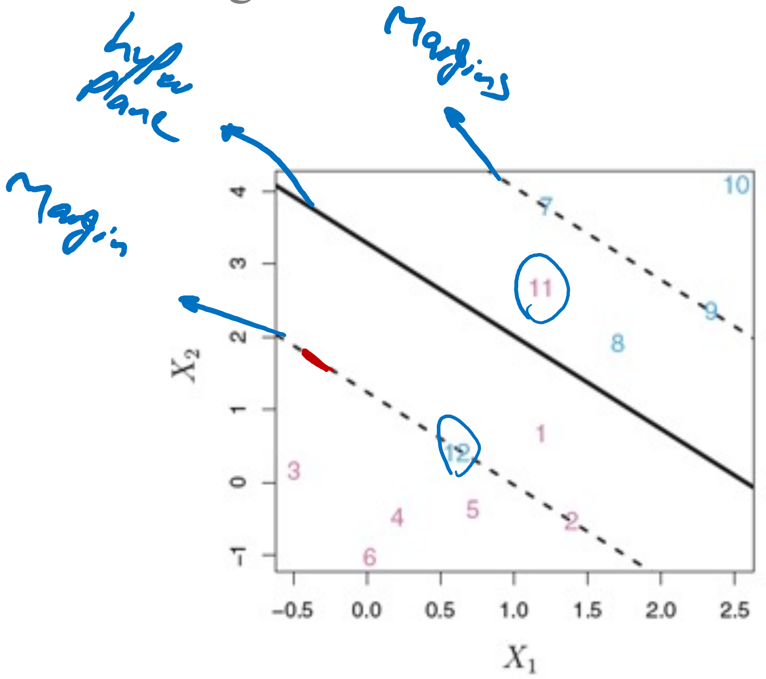


wrong side of the margin
wrong side of margin but right side of hyp

Support Vector Classifier (SVC) – Soft Margin

- **Solution**: we can extend the concept of a separating hyperplane in order to develop a hyperplane that **almost** separates the classes, using a so-called **soft margin**.
- The generalization of the maximal margin classifier using soft margin is known as the support vector classifier (SVC).
- It could be worthwhile to misclassify a few training observations in order to do a better job in classifying the remaining observations.





1 wrong side of margin
8 wrong " " "
11 wrong " " hyperplane
12 " " " "

SVC optimization problem

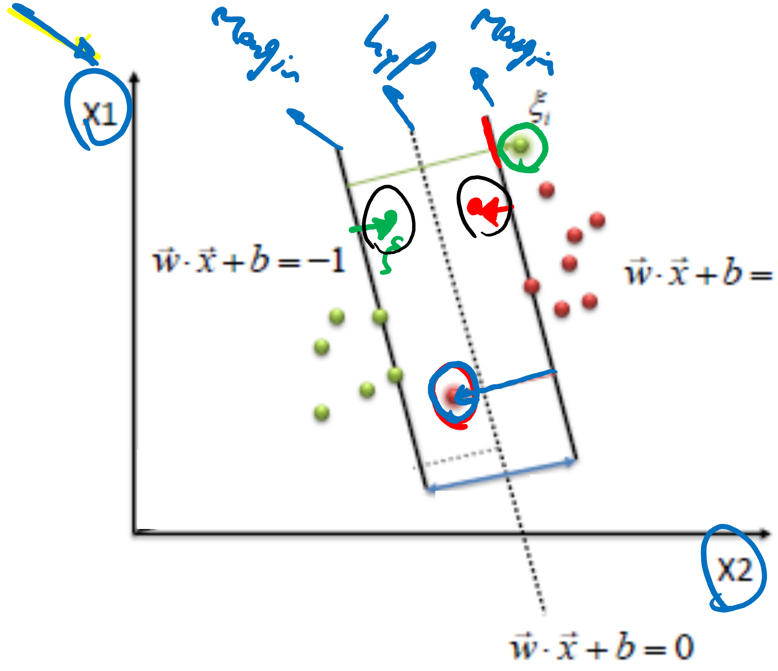
- Soft margin classification adds a penalty (C) to the objective function for observations in the training set that are misclassified. In essence, the SVM algorithm will choose a decision boundary that optimizes the trade-off between a wider margin and a lower total error penalty.
- Slack variable ξ_i allow some observations to fall on the wrong side of the margin, but will penalized them by parameter C: Cost of misclassification

Soft Margin

$$\underset{w,b}{\text{Min}} \ \frac{1}{2}||w||^2 + C\sum_{i=1}^{I}\xi_i$$

$$s.t. \ \ y_i\left(\sum_{k=1}^{K} w_k x_{i,k} + b\right) \geq 1 - \xi_i \qquad \xi_i \geq 0 \ \forall_i$$

$$\begin{cases} \sum_{k=1}^{K} w_k x_{i,k} + b \geq +1 - \xi_i & \text{when } y_i = +1 \\ \sum_{k=1}^{K} w_k x_{i,k} + b \leq -1 + \xi_i & \text{when } y_i = -1 \end{cases}$$



Regularization parameter

SVC → $\min \frac{1}{2}\|w\|^2 + C \sum \xi_i$

C: Cost of misclassification

Small C :
wide margin : high bias : low variance

Large C: C↑
narrow margin : low bias : High variance

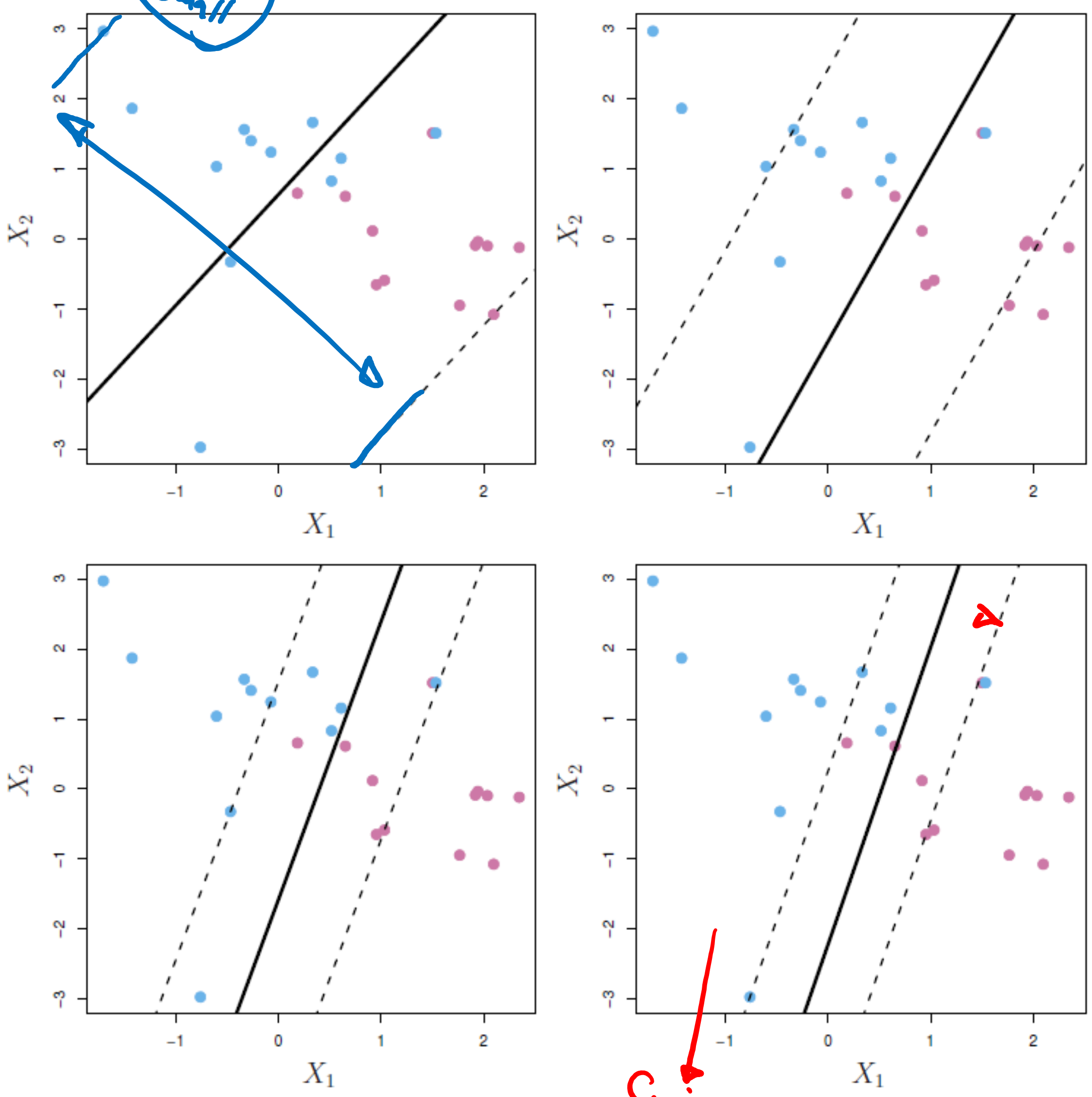
Part 5

MMC → SVC → SVM

Soft + kernel trick

C is small

wide



Narrow

C is Large

MMC solution

$$\underset{\mathbf{w},b}{\text{argmin}}\ \frac{1}{2}||\mathbf{w}||^2 \ \text{ s.t. } y_i\left(\sum_{k=1}^{K} w_k x_{i,k} + b\right) \geq 1$$

$$L(\mathbf{w}, b, \boldsymbol{\lambda}) = \frac{1}{2}||\mathbf{w}||^2 + \sum_{i=1}^{I} \lambda_i \left(y_i\left(\sum_{k=1}^{K} w_k x_{i,k} + b\right) - 1\right)$$

$$\frac{\partial L}{\partial \mathbf{w}} L(\mathbf{w}, b, \boldsymbol{\lambda}) = \mathbf{0}, \quad \frac{\partial L}{\partial b} L(\mathbf{w}, b, \boldsymbol{\lambda}) = 0,$$

$$\mathbf{w}^* = \sum_{i=1}^{I} \lambda_i u_i \mathbf{x}_i.$$