Statistical Learning

https://github.com/ggorr/Machine-Learning/tree/master/ISLR

9
Support Vector Machines

- 9.1 Maximal Margin Classifier
- 9.2 Support Vector Classifiers
- 9.3 Support Vector Machines
- 9.4 SVMs with More than Two Classes
- 9.5 Relationship to Logistic Regression
- 9.6 Lab: Support Vector Machines
- 9.7 Exercises

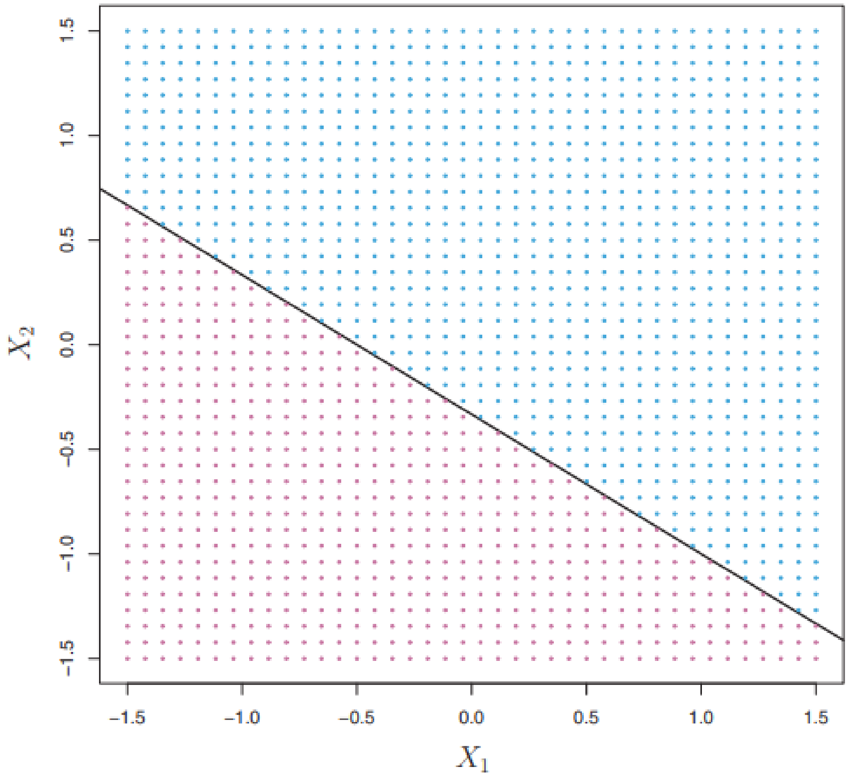
9.1 Maximal Margin Classifier

- 9.1.1 What Is a Hyperplane?
- 9.1.2 Classification Using a Separating Hyperplane
- 9.1.3 The Maximal Margin Classifier
- 9.1.4 Construction of the Maximal Margin Classifier
- 9.1.5 The Non-separable Case

9.1.1 What Is a Hyperplane?

- A hyperplane in p-dimensional space
  - A flat affine subspace of dimension $p-1$
- 2-dimension
  - $\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$
  - A line
- p-dimension
  - $\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p = 0$

- A hyperplane divides p-dimensional space into two halves
  - $\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p > 0$
  - $\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p < 0$

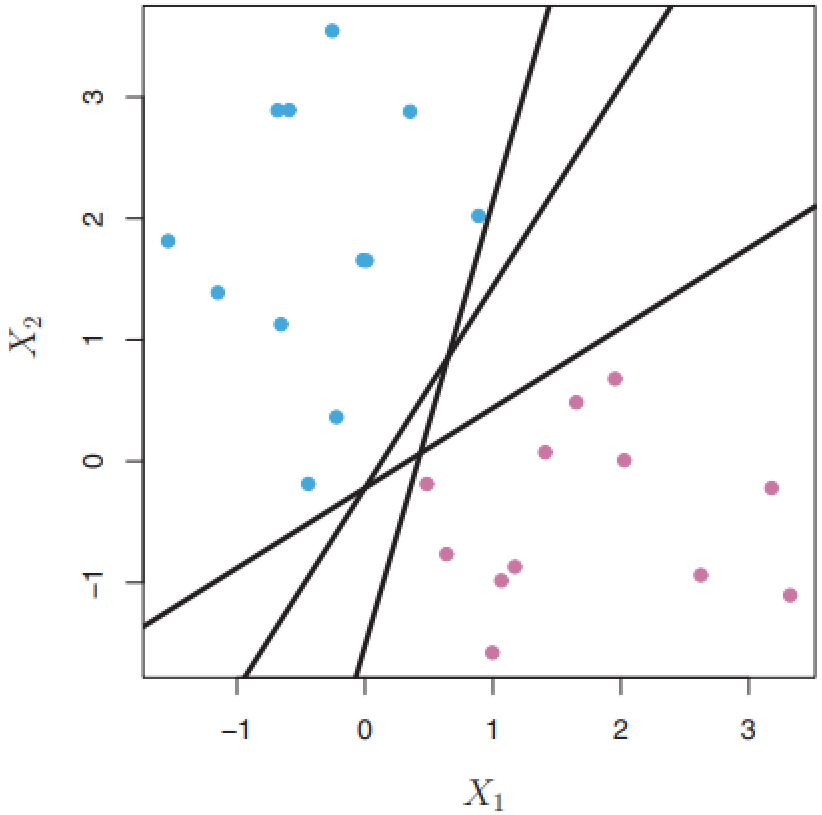


9.1.2 Classification Using a Separating Hyperplane

- Training observations
  - $x_1 = \begin{bmatrix} x_{11} \\ \vdots \\ x_{1p} \end{bmatrix}, \dots, x_n = \begin{bmatrix} x_{n1} \\ \vdots \\ x_{np} \end{bmatrix}$
  - $y_1, \dots, y_n \in \{-1, 1\}$
- Test observation
  - $x^* = \begin{bmatrix} x_1^* \\ \vdots \\ x_p^* \end{bmatrix}$
  - $y^* = ?$

Separating hyperplane

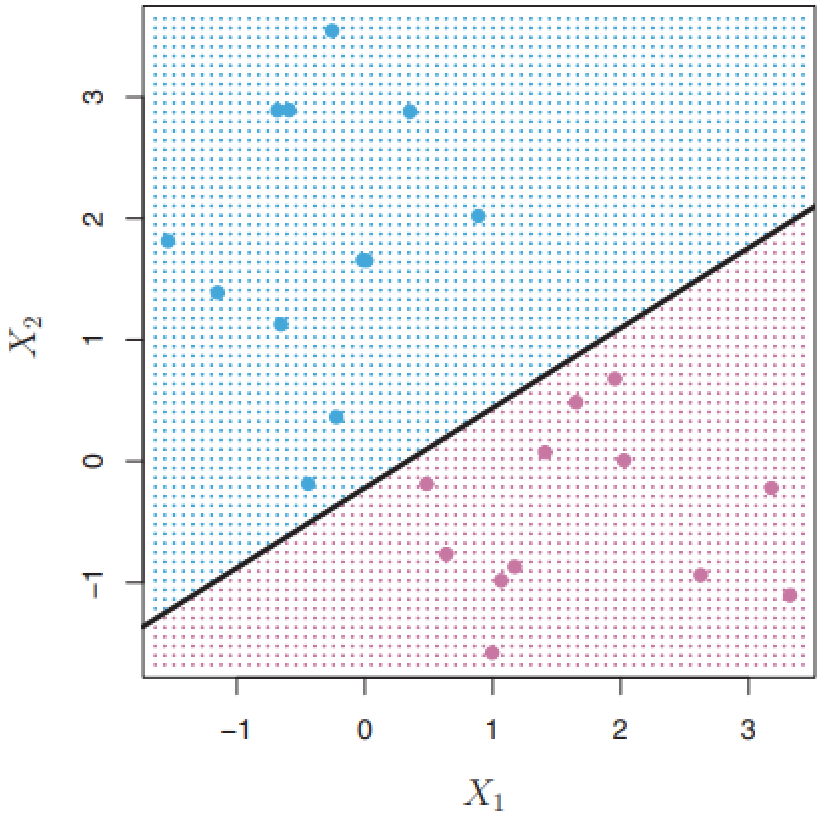
- Separating hyperplane
  - A hyperplane $\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p = 0$ satisfies
    $$\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} > 0 \text{ if } y_i = 1$$
    $$\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} < 0 \text{ if } y_i = -1$$
    for all $i = 1, \ldots, n$
  - Equivalently,
    $$y_i(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}) > 0$$
    for all $i = 1, \ldots, n$



- Let

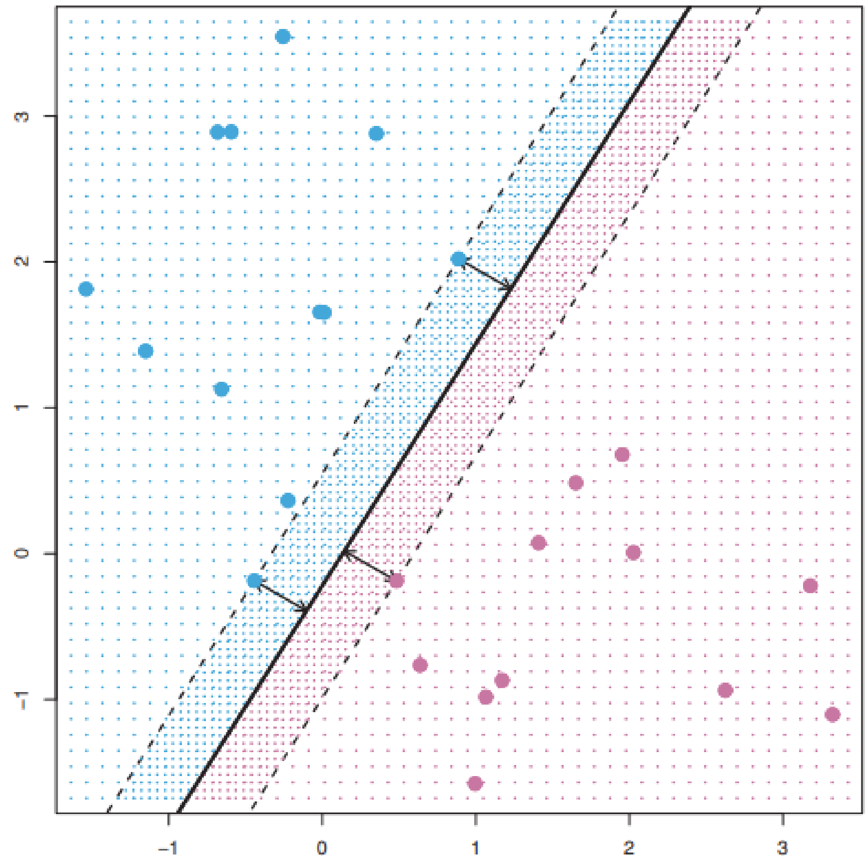
$$f(x^*) = \beta_0 + \beta_1 x_1^* + \cdots + \beta_p x_p^*$$

- $y^* = \begin{cases} 1 & \text{if } f(x^*) > 0 \\ -1 & \text{if } f(x^*) > 0 \end{cases}$



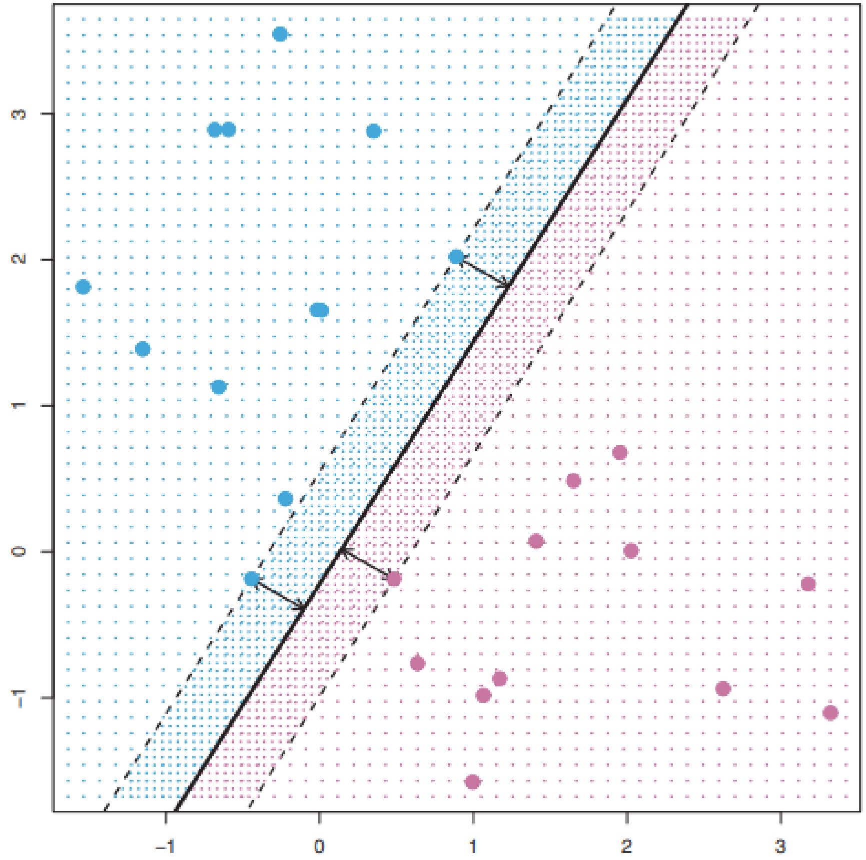
9.1.3 The Maximal Margin Classifier

- Margin
  - The minimal perpendicular distance from the observations to the separating hyperplane
- Maximal margin hyperplane (a.k.a. optimal separating hyperplane)
  - The separating hyperplane that is farthest from the training observations
- Maximal margin classifier
  - To classify test observation using maximal margin hyperplane



Support vector

- Support vector
  - training observations that are equidistant from the maximal margin hyperplane
- Hyperplane in the right figure has 3 support vectors



9.1.4 Construction of the Maximal Margin Classifier

- Optimization problem:

$$\underset{\beta_0,\ldots,\beta_p}{\text{maximize}}\, M$$

subject to

$$\sum_{j=1}^{p} \beta_j^2 = 1$$

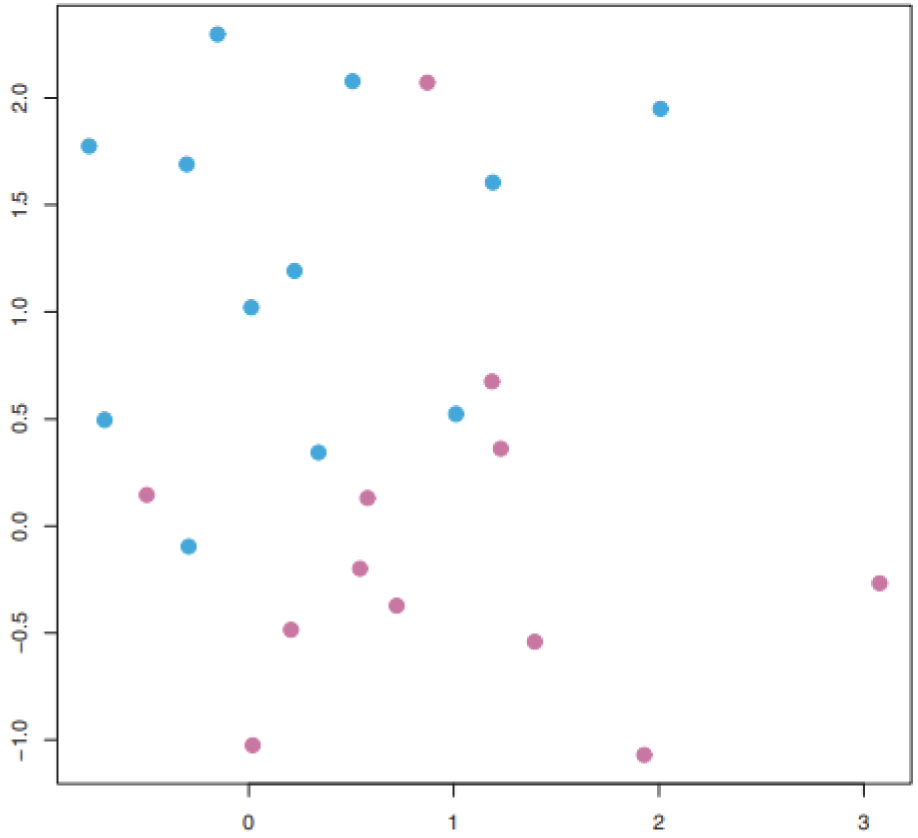
and

$$y_i(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}) > M$$

for all $i = 1, \ldots, n$

9.1.5 The Non-separable Case

- Support vector classifier.
  - Find a hyperplane that almost separates the classes
  - Using a soft margin

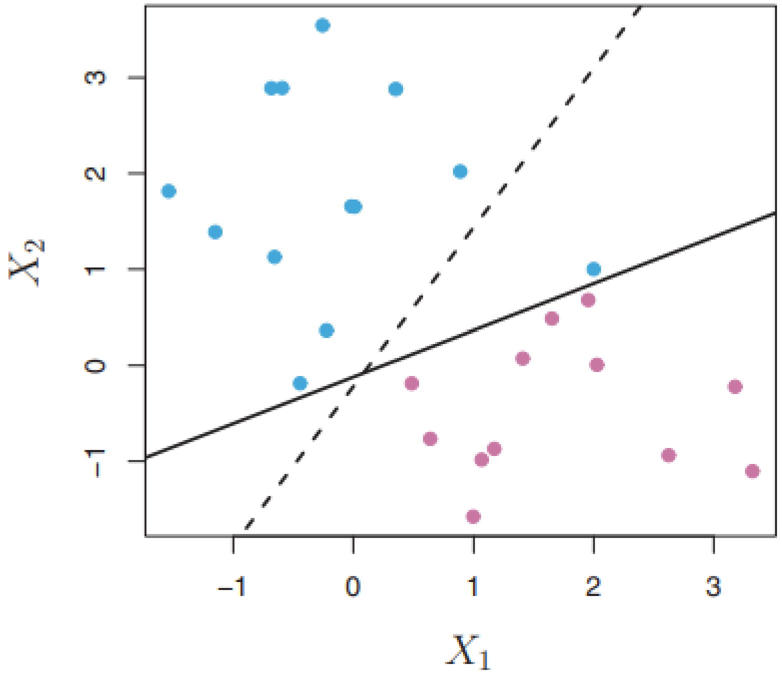


9.2 Support Vector Classifiers

- 9.2.1 Overview of the Support Vector Classifier
- 9.2.2 Details of the Support Vector Classifier

9.2.1 Overview of the Support Vector Classifier

- the maximal margin hyperplane
  - Sometimes, it has only a tiny margin
  - Extremely sensitive to a change in a single observation
  - Sensitive to error



support vector classifier

- support vector classifier
  = soft margin classifier
  - Based on a hyperplane that may not perfectly separate the two classes
  - Greater robustness to individual observations
  - Better classification of most of the training observations

9.2.2 Details of the Support Vector Classifier

- Optimization problem:

$$\underset{\beta_0,\dots,\beta_p,\epsilon_1,\dots,\epsilon_n}{\text{maximize}} \quad M$$

subject to

$$\sum_{j=1}^{p} \beta_j^2 = 1$$

and

$$y_i\left(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\right) > M(1 - \epsilon_i),$$

$$\epsilon_i \geq 0,\ \sum_{i=1}^{n} \epsilon_i \leq C$$

for all $i = 1, \dots, n$ where $C \geq 0$