

12.1 Figure 12.12 shows about 20 observations with two x-variables. The observations belong to one of two classes: triangle (red) or square (blue). Also shown is the separating line as well as the two margin lines.

- Identify all support vectors on the graph by putting a small "s" next to each observation.
- Identify all observations that are misclassified by drawing a circle around misclassified observations.
- The graph shows something unusual. Can you find it? *Unsure*

12.2 Consider the toy data in Table 12.2.

- By hand or using software, create a scatter plot of the data. Label the points with their classes.
- Sketch the optimal separating hyperplane. Give the equation for the hyperplane. Give an expression for which the points are classified as yellow. Hint: The optimal hyperplane usually has three support vectors, two on one side and one of the other. (One exception is the subject of question 12.3). If you see several potential optimal hyperplanes, explore each and choose the one with the greatest margin M .
- On the same graph, sketch a hyperplane that is not optimal.
- On the same graph, add an additional observation such that the two classes are still separable, but no longer linearly separable.

Table 12.2 Toy data for a support vector classifier

x_1	x_2	y
0	4	yellow
3	1	yellow
2	3	yellow
4	4	green
3	5	green
5	6	green
6	4	green

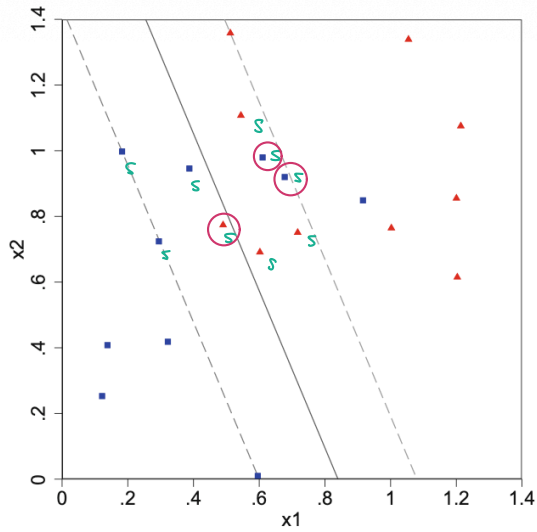


Fig. 12.12 Linear SVM with two x-variables with separating line and margin lines

① Give equation for the hyperplane

$$B_0 + B_1 + \dots + B_p x_p = 0$$

Potential support vectors: $(3, 1), (4, 4), (2, 3)$

$$y_i (B_0 + B_1 x_i + B_2 x_{i2}) = 1$$

$$(2, 3) \rightarrow -(B_0 + B_1(2) + B_2(3)) = 1$$

$$(4, 4) \rightarrow B_0 + B_1(4) + B_2(4) = 1$$

$$(3, 1) \rightarrow -(B_0 + B_1(3) + B_2(1)) = 1$$

To maximize margin, minimize $\|B\|$

$$\text{The margins must be } B_0 + B'x = \pm 1$$

$$B_1 + B'x = -1$$

$$\begin{aligned} \text{Distance between two parallel lines} &\Rightarrow d = \frac{|C_2 - C_1|}{\sqrt{a^2 + b^2}} \\ &= \frac{|+1 - (-1)|}{\sqrt{a^2 + b^2}} \\ &= \frac{2}{\|B\|} \end{aligned}$$

②

$$\text{① eliminate } B_0. \quad B_0 + 4B_1 + 4B_2 = 1$$

$$-(B_0 - 2B_1 - 3B_2) = -1$$

② Solve for B_2

$$2B_1 + 1B_2 = 2$$

$$1B_1 + 3B_2 = 2$$

$$\hookrightarrow B_1 = 2 - 3B_2$$

$$2(2 - 3B_2) + 1B_2 = 2$$

$$4 - 6B_2 + 1B_2 = 2$$

$$-5B_2 = -2$$

$$\begin{aligned} 2B_1 + \frac{2}{5} &= 2 \\ 2B_1 &= \frac{10}{5} - \frac{2}{5} \\ B_1 &= \frac{8}{5} = \frac{2}{10} \\ B_2 &= \frac{2}{5} \end{aligned}$$

③

$$B_0 + 4B_1 + 4B_2 = 1$$

$$-(B_0 + 3B_1 + 1B_2) = -1$$

$$1B_1 + 3B_2 = 2$$

Hence, the best line becomes $-\frac{19}{5} + \frac{9}{5}x_1 + \frac{2}{5}x_2 = 0$

12.4 Consider a support vector classification of two classes with a linear kernel. The parameter C allows for a certain amount of error. For the C -values specified below, what is the maximum number of misclassified observations in the training data? (Observations on the boundary are not inside the margin.) Explain briefly. If the training data cannot misclassify any observations, are there any observations inside the margin? Explain briefly.

- (a) $C = 10$
- (b) $C = 0.5$
- (c) $C = 0$

$$12.4 \quad y_i \left(\beta_0 + \sum_{j=1}^p \beta_{j1} x_{ij} + \sum_{j=1}^p \beta_{j2} x_{ij}^2 \right) \geq M(1 - \epsilon_i)$$

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

Prediction $\hat{f}(x) = \hat{\beta}_0 + \sum_{i \in S} \hat{\alpha}_i y_i \langle x, x_i \rangle$

Complementary slackness $\alpha[y_i(\beta_0 + x' \beta) - 1] = 0$

- Ⓐ $\sum_{i=1}^n \epsilon_i \leq C$
 - $\epsilon > 1 \rightarrow$ misclassified
 - $\epsilon = 0$ outside the margin
 - $0 < \epsilon < 1 \rightarrow$ point is inside margin but correct side

max classified = 101

1 misclassification is 1 slack.

Hence, if $C = 10$, 10 misclassification.

- Ⓑ $C = 0.5$, since $\epsilon > 1$ is classified as misclassification

0.5 is not a misclassification. Hence D.

- Ⓒ Same reasoning as Ⓑ

12.5 The two graphs in Fig. 12.13 show two linear SVM classifiers for the same data. Which of the two had the higher value of the tuning parameter C ? Explain briefly.

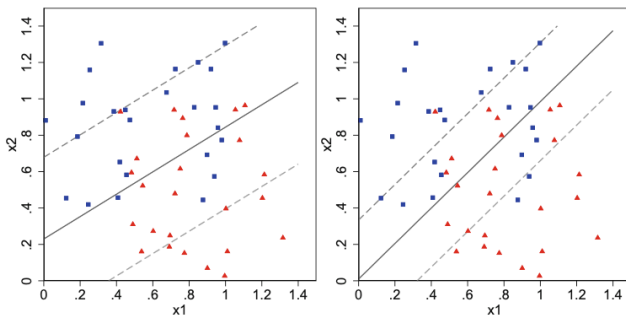


Fig. 12.13 Two linear SVMs for two x-variables with different costs

12.5

Right graph has higher C .

Higher $C \rightarrow$ more violations tolerated $\rightarrow$ more tries harder to classify correctly $\rightarrow$ Narrow margin

12.6 For support vector regression, what role do the parameters specified below play in terms of the variance/bias tradeoff? Explain briefly.

- (a) Parameter ϵ
- (b) Parameter C
- (c) Parameter γ

12.6 Ⓐ ϵ = Defines the margin tolerance / slack variables.

Since $\epsilon_i > 0$ and $\sum_{i=1}^n \epsilon_i \leq 0$, larger ϵ allows for more points ignored

Ⓑ C = penalty for points outside the tube (violations)

Ⓒ γ = Determines how fast the influence of neighboring observations on the classification of the

test observation drops off as a function of distance. Large γ leads to a local fit

→ How far the influence of a single training points reached, low bias and high variance