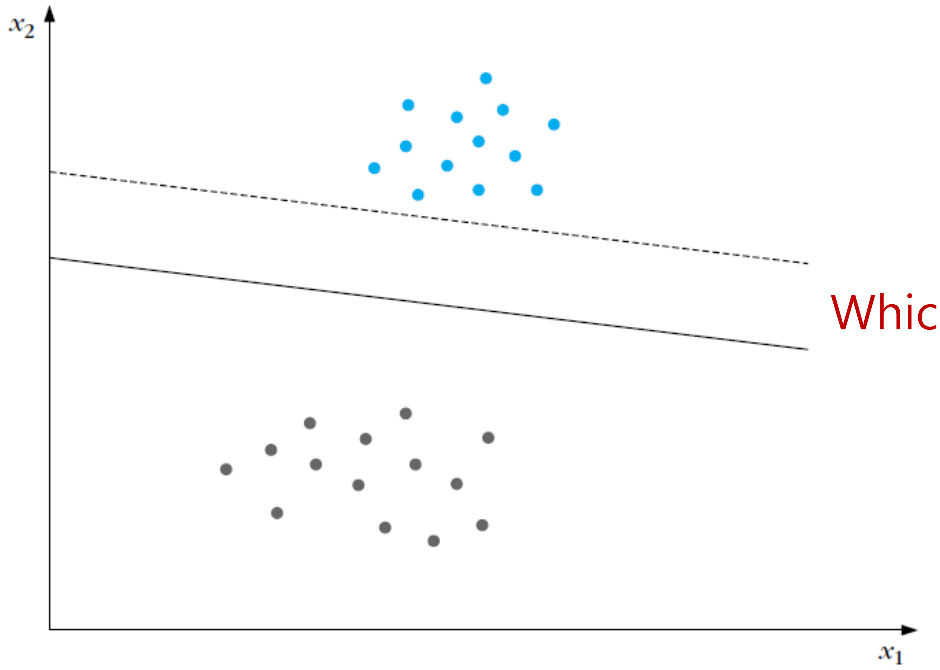
SUPPORT VECTOR MACHINES

Separable Classes

- The goal is again to design a separating hyperplane

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0$$



Which line is better?

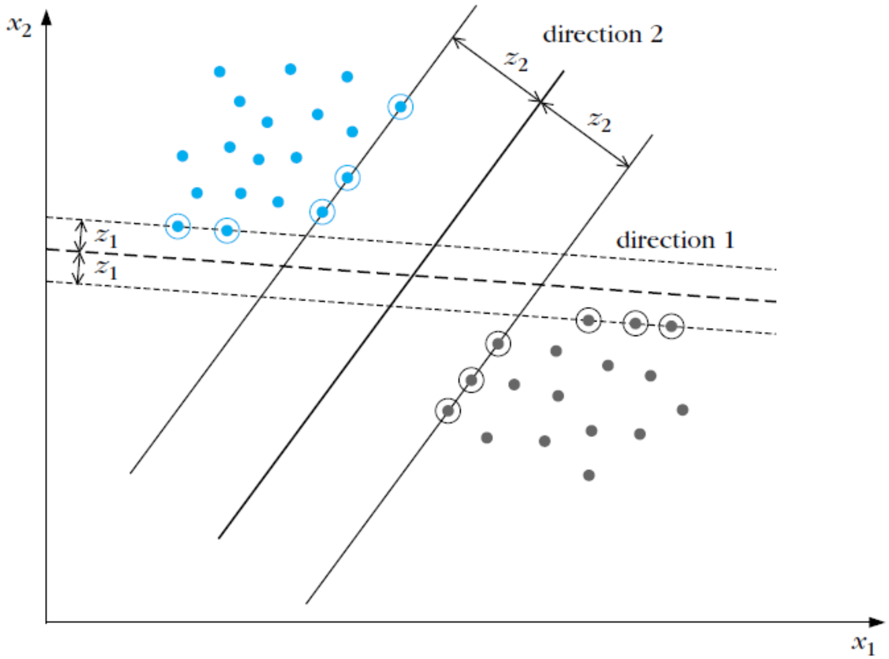
Separable Classes

- Generalization performance of a classifier: capability to operate accurately on non-training data

- In this figure
  - Margin: $2z_1$ or $2z_2$, which is given by

    $$\frac{1}{\|\mathbf{w}\|}+\frac{1}{\|\mathbf{w}\|}=\frac{2}{\|\mathbf{w}\|}$$
  - while

    $$\mathbf{w}^T\mathbf{x}+w_0 \geq 1, \quad \forall\mathbf{x}\in\omega_1$$
    $$\mathbf{w}^T\mathbf{x}+w_0 \leq -1, \forall\mathbf{x}\in\omega_2$$



Separable Classes

- The SVM problem can be formulated as computing the parameters $\mathbf{w}$, w_0 so that

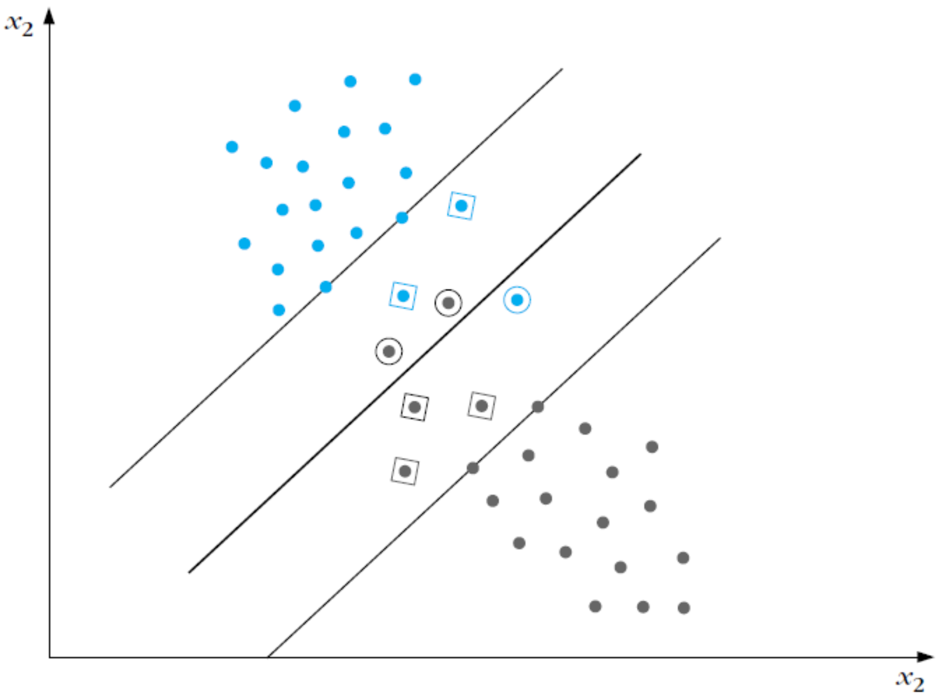
$$\text{minimize } J(\mathbf{w}, w_0) \equiv \frac{1}{2} ||\mathbf{w}||^2$$
$$\text{subject to } y_i(\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1, \quad i = 1,2,\dots,N$$
$$\text{where } y_i = 1 \text{ if } \mathbf{x}_i \in \omega_1 \text{ and } 0 \text{ if } \mathbf{x}_i \in \omega_2$$

- This is a standard optimization problem and can be solved using, for example, Matlab
- The optimization techniques are beyond the scope of this lecture

Non-separable Classes

- The margin is defined as the distance between the parallel hyperplanes $\mathbf{w}^T\mathbf{x} + w_0 = \pm 1$

- The training vectors belong to one of the following three categories:
  1) Vectors that fall outside the band and are correctly classified
     $$y_i(\mathbf{w}^T\mathbf{x} + w_0) > 1$$
  2) Vectors falling inside the band and are correctly classified
     $$0 \leq y_i(\mathbf{w}^T\mathbf{x} + w_0) < 1$$
  3) Vectors that are misclassified
     $$y_i(\mathbf{w}^T\mathbf{x} + w_0) < 0$$



Non-separable Classes

- Constraints

$$y_i[\mathbf{w}^T\mathbf{x} + w_0] \geq 1 - \xi_i$$

with slack variable ξ_i

1) $\xi_i = 0$
2) $0 < \xi_i \leq 1$
3) $\xi_i > 1$

- Cost function

$$J(\mathbf{w}, w_0, \xi) \equiv \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} I(\xi_i)$$

where

$$I(\xi_i) = \begin{cases} 1 & \xi_i > 0 \\ 0 & \xi_i = 0 \end{cases}$$

- Because $I(\xi_i)$ is not differentiable, this is not easy to optimize

Non-separable Classes

- Problem formulation

Minimize $J(\mathbf{w}, w_0, \xi) \equiv \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i$

Subject to $y_i[\mathbf{w}^T\mathbf{x}_i + w_0] \geq 1 - \xi_i,\ i = 1,2, \dots, N$

$\xi_i \geq 0,\ i = 1,2, \dots, N$

- This is a typical optimization problem, which can be easily solved

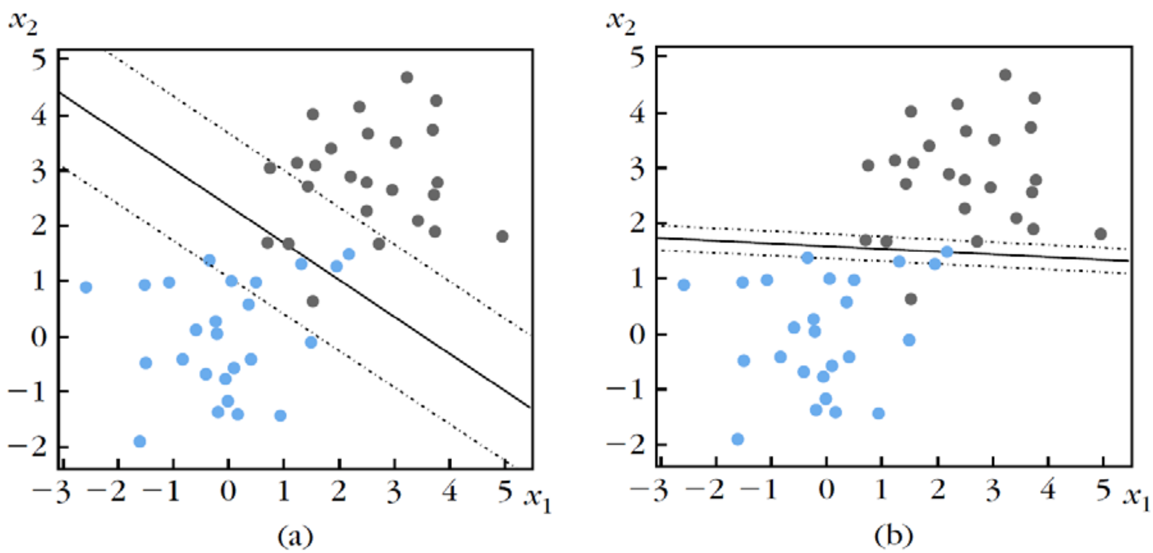


FIGURE 3.13

An example of two nonseparable classes and the resulting SVM linear classifier (full line) with the associated margin (dotted lines) for the values (a) $C = 0.2$ and (b) $C = 1000$. In the latter