

Figure 1.24 Schematic illustration of the joint probabilities $p(x, \mathcal{C}_k)$ for each of two classes plotted against x , together with the decision boundary $x = \hat{x}$. Values of $x \geq \hat{x}$ are classified as class \mathcal{C}_2 and hence belong to decision region \mathcal{R}_2 , whereas points $x < \hat{x}$ are classified as \mathcal{C}_1 and belong to \mathcal{R}_1 . Errors arise from the blue, green, and red regions, so that for $x < \hat{x}$ the errors are due to points from class \mathcal{C}_2 being misclassified as \mathcal{C}_1 (represented by the sum of the red and green regions), and conversely for points in the region $x \geq \hat{x}$ the errors are due to points from class \mathcal{C}_1 being misclassified as \mathcal{C}_2 (represented by the blue region). As we vary the location \hat{x} of the decision boundary, the combined areas of the blue and green regions remains constant, whereas the size of the red region varies. The optimal choice for \hat{x} is where the curves for $p(x, \mathcal{C}_1)$ and $p(x, \mathcal{C}_2)$ cross, corresponding to $\hat{x} = x_0$, because in this case the red region disappears. This is equivalent to the minimum misclassification rate decision rule, which assigns each value of x to the class having the higher posterior probability $p(\mathcal{C}_k|x)$.

	cancer	normal
cancer	0	1000
normal	1	0

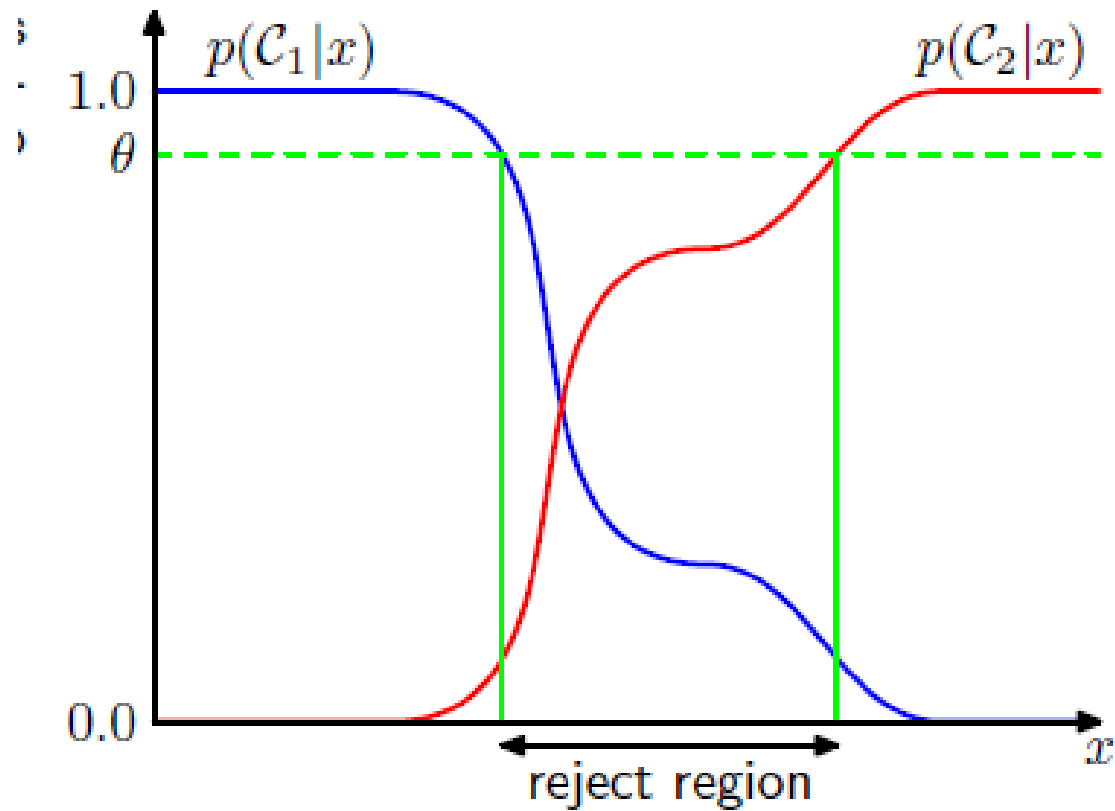
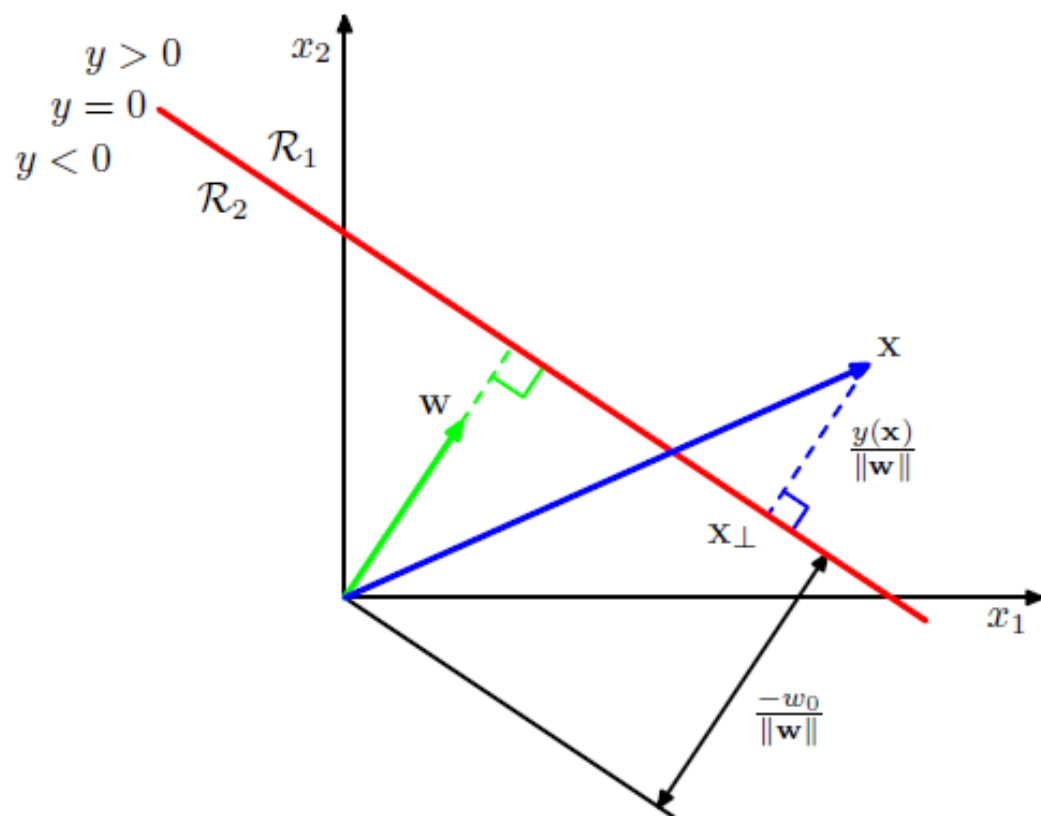


Figure 1.26 Illustration of the reject option. Inputs x such that the larger of the two posterior probabilities is less than or equal to some threshold θ will be rejected.

Figure 4.1 Illustration of the geometry of a linear discriminant function in two dimensions. The decision surface, shown in red, is perpendicular to \mathbf{w} , and its displacement from the origin is controlled by the bias parameter w_0 . Also, the signed orthogonal distance of a general point \mathbf{x} from the decision surface is given by $y(\mathbf{x})/\|\mathbf{w}\|$.



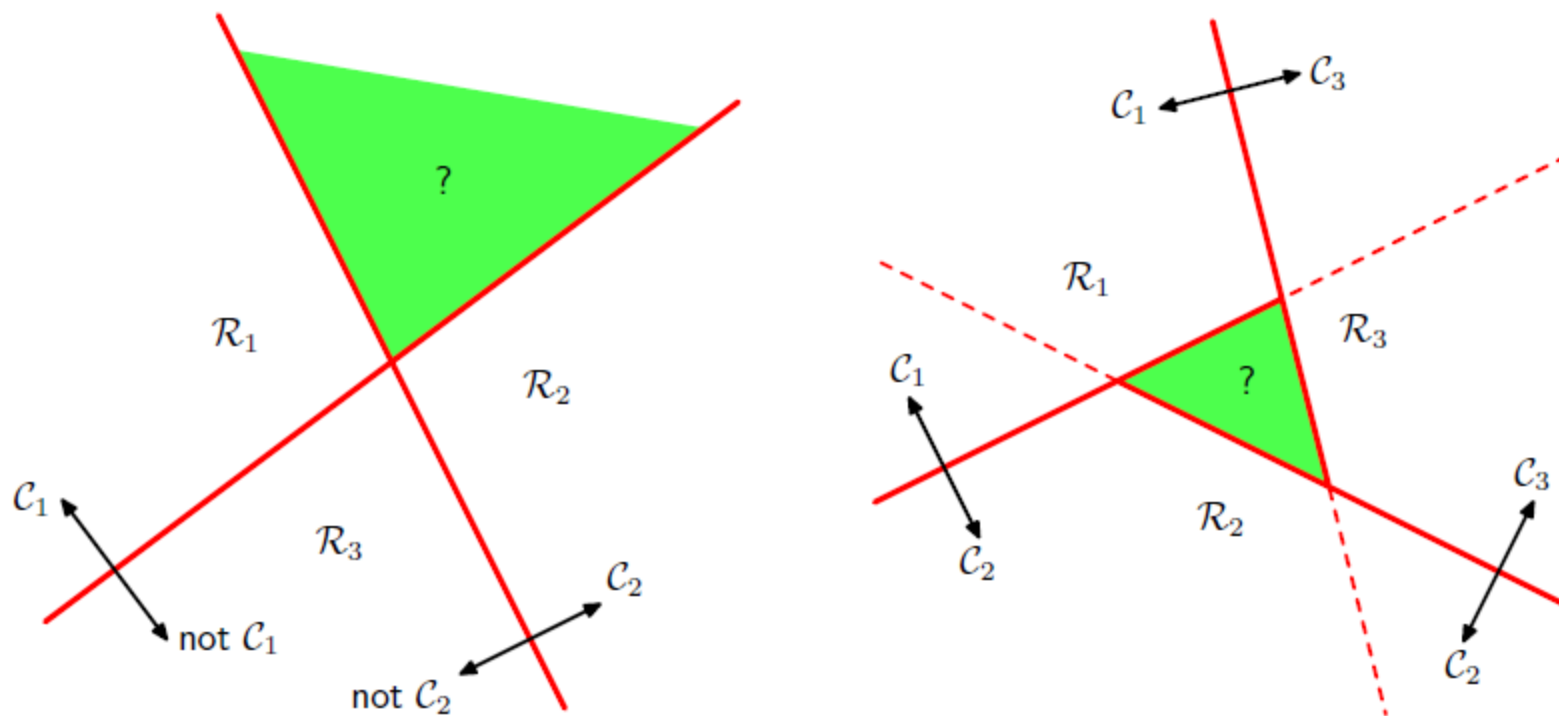
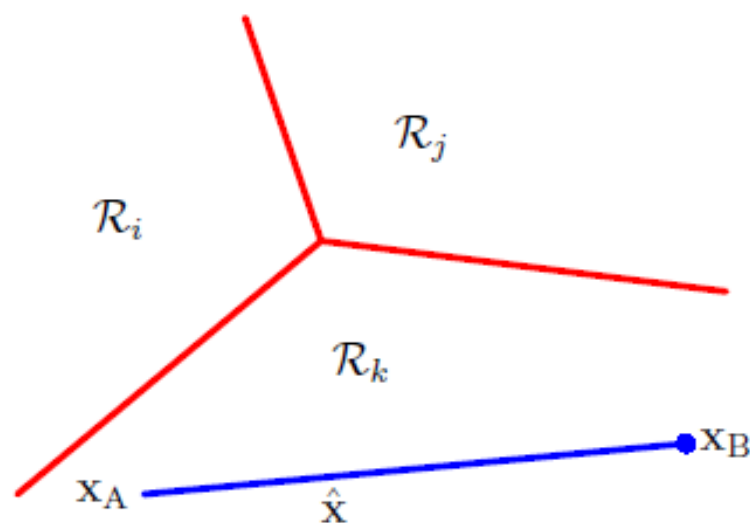


Figure 4.2 Attempting to construct a K class discriminant from a set of two class discriminants leads to ambiguous regions, shown in green. On the left is an example involving the use of two discriminants designed to distinguish points in class \mathcal{C}_k from points not in class \mathcal{C}_k . On the right is an example involving three discriminant functions each of which is used to separate a pair of classes \mathcal{C}_k and \mathcal{C}_j .

Figure 4.3 Illustration of the decision regions for a multiclass linear discriminant, with the decision boundaries shown in red. If two points x_A and x_B both lie inside the same decision region \mathcal{R}_k , then any point \hat{x} that lies on the line connecting these two points must also lie in \mathcal{R}_k , and hence the decision region must be singly connected and convex.



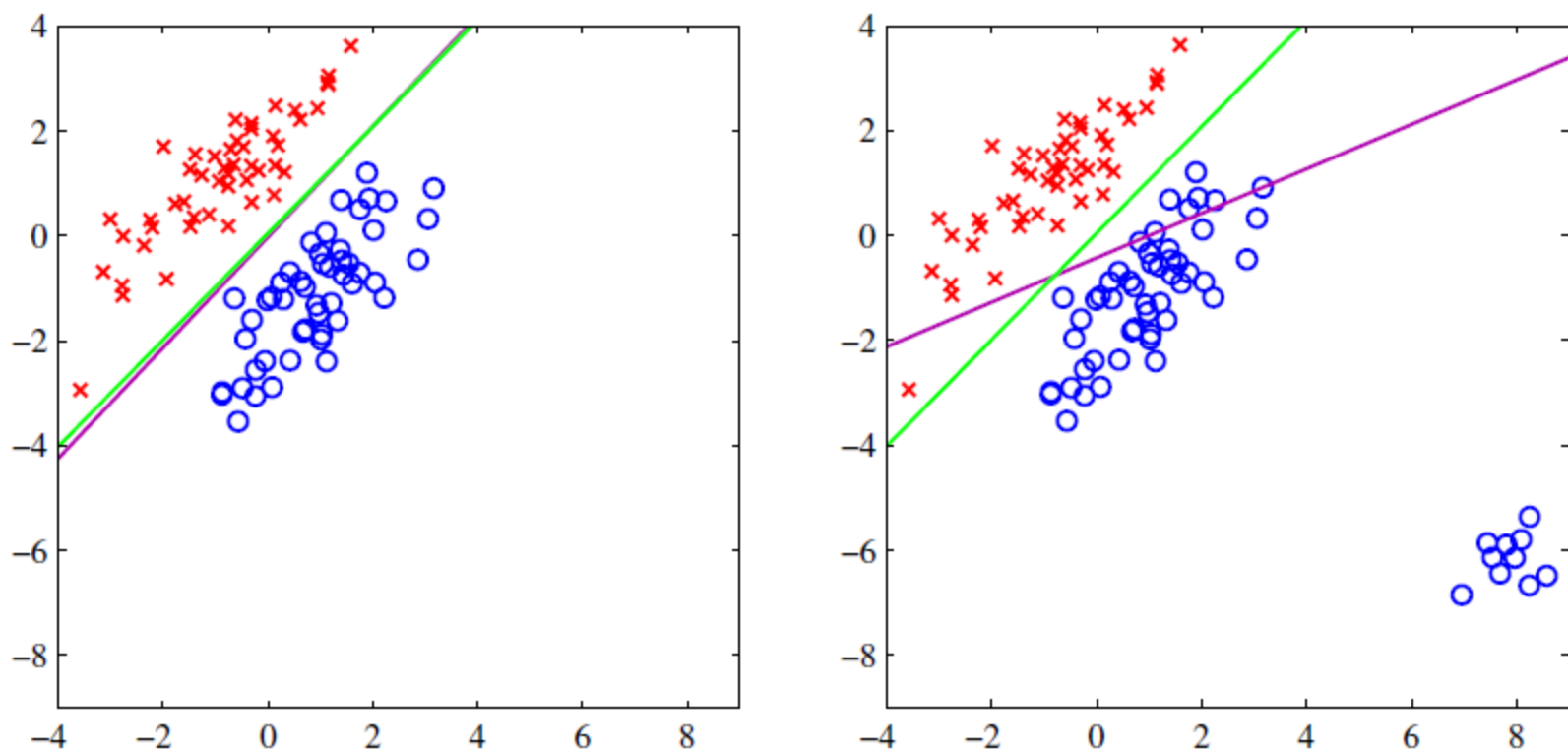


Figure 4.4 The left plot shows data from two classes, denoted by red crosses and blue circles, together with the decision boundary found by least squares (magenta curve) and also by the logistic regression model (green curve), which is discussed later in Section 4.3.2. The right-hand plot shows the corresponding results obtained when extra data points are added at the bottom left of the diagram, showing that least squares is highly sensitive to outliers, unlike logistic regression.

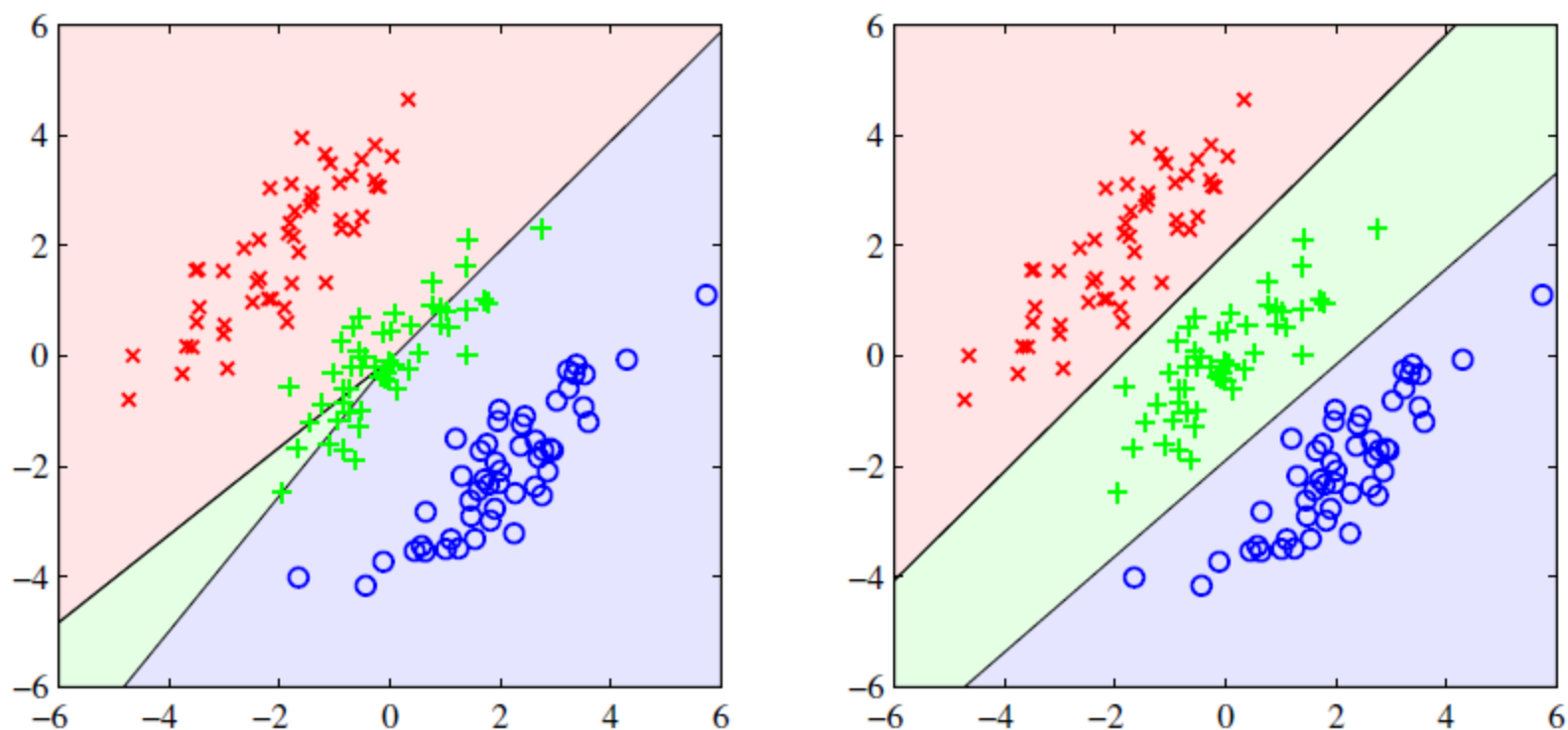


Figure 4.5 Example of a synthetic data set comprising three classes, with training data points denoted in red (\times), green ($+$), and blue (\circ). Lines denote the decision boundaries, and the background colours denote the respective classes of the decision regions. On the left is the result of using a least-squares discriminant. We see that the region of input space assigned to the green class is too small and so most of the points from this class are misclassified. On the right is the result of using logistic regressions as described in Section 4.3.2 showing correct classification of the training data.

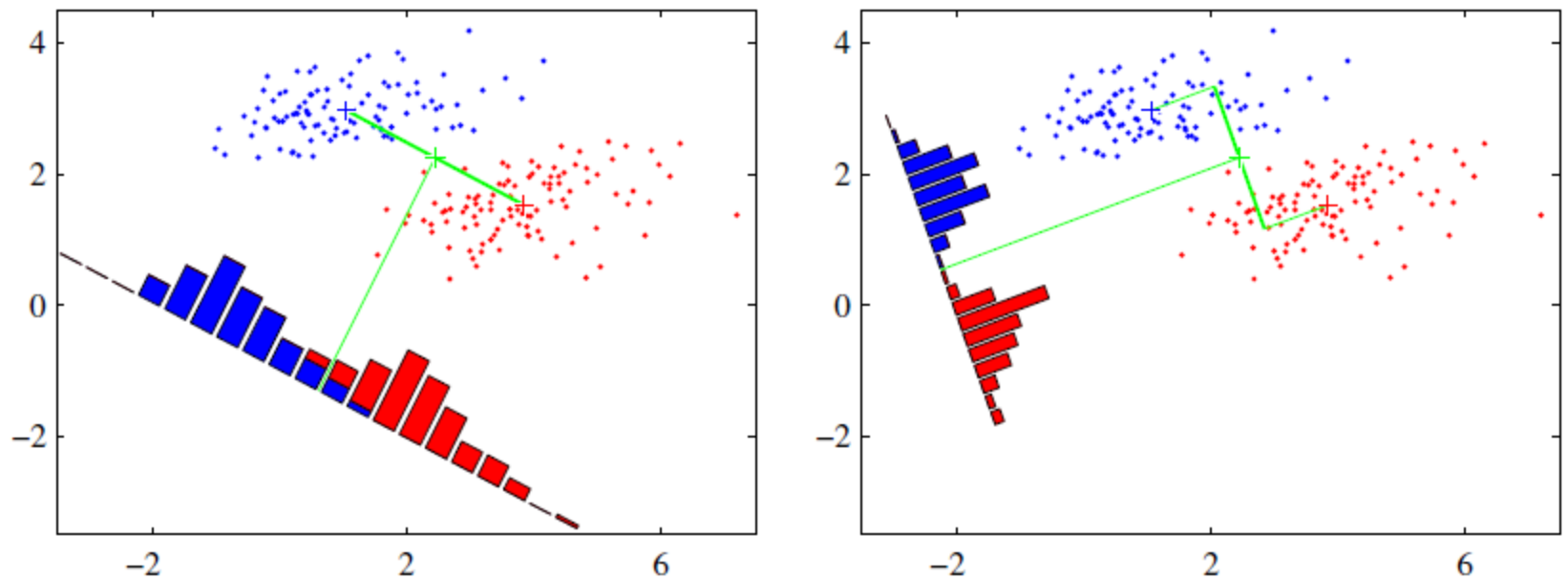


Figure 4.6 The left plot shows samples from two classes (depicted in red and blue) along with the histograms resulting from projection onto the line joining the class means. Note that there is considerable class overlap in the projected space. The right plot shows the corresponding projection based on the Fisher linear discriminant, showing the greatly improved class separation.

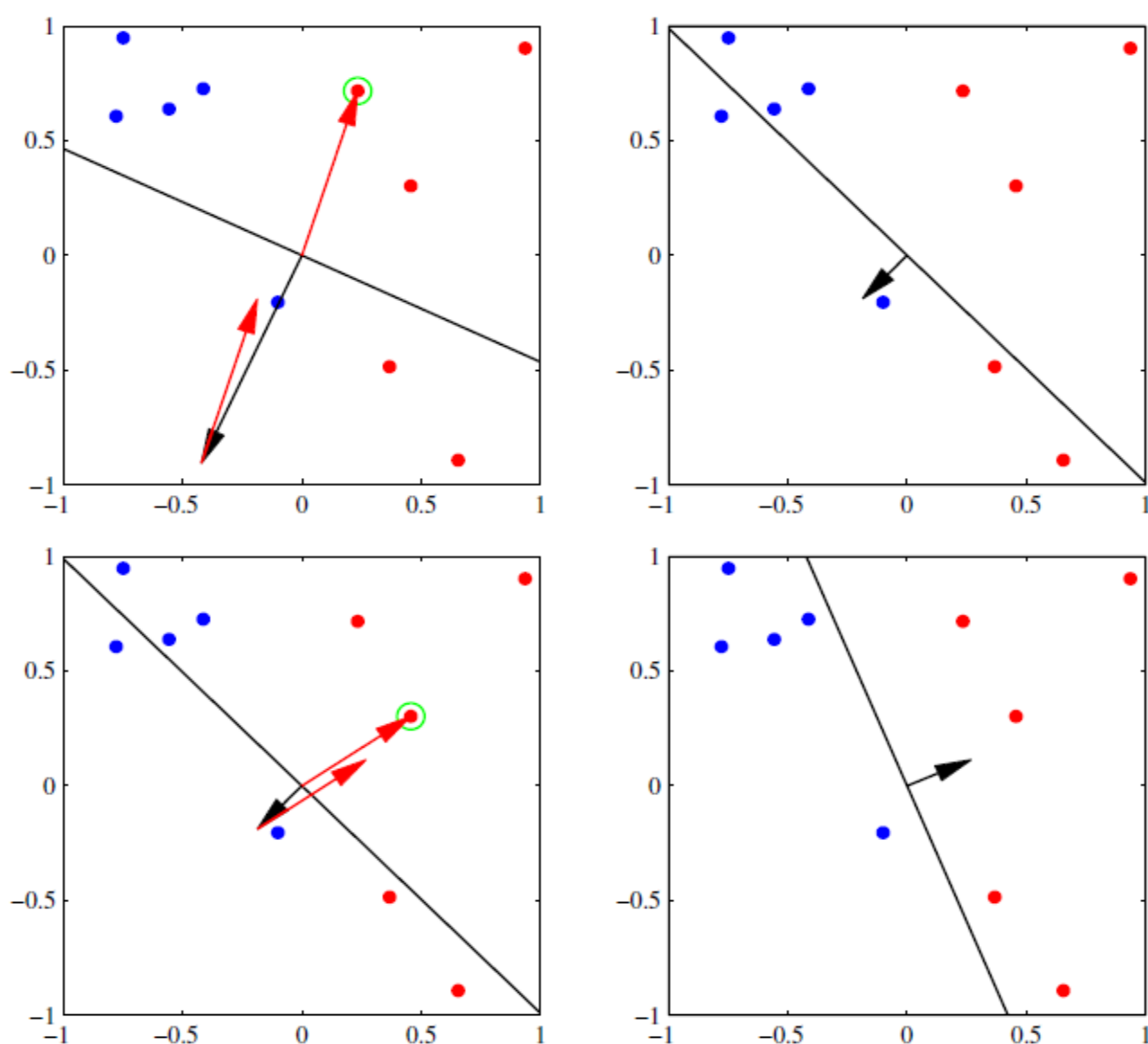


Figure 4.7 Illustration of the convergence of the perceptron learning algorithm, showing data points from two classes (red and blue) in a two-dimensional feature space (ϕ_1, ϕ_2) . The top left plot shows the initial parameter vector \mathbf{w} shown as a black arrow together with the corresponding decision boundary (black line), in which the arrow points towards the decision region which classified as belonging to the red class. The data point circled in green is misclassified and so its feature vector is added to the current weight vector, giving the new decision boundary shown in the top right plot. The bottom left plot shows the next misclassified point to be considered, indicated by the green circle, and its feature vector is again added to the weight vector giving the decision boundary shown in the bottom right plot for which all data points are correctly classified.