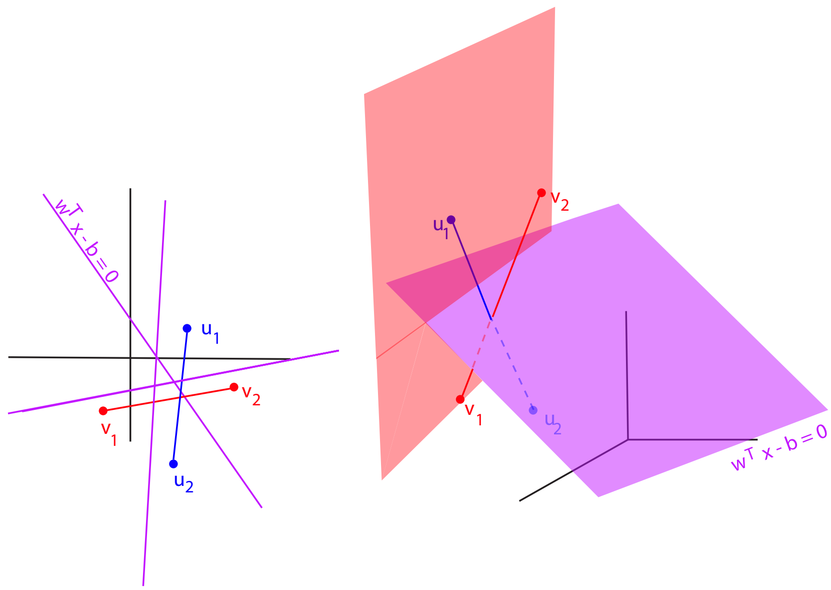

Figure 50.12: Two examples in which it is impossible to find purple hyperplanes which separate the red and blue points.

Remark: Write $m = p + q$. The reader should be aware that in machine learning the classification problem is usually defined as follows. We assign m so-called *class labels* $y_k = \pm 1$ to the data points in such a way that $y_i = +1$ for each blue point u_i, and $y_{p+j} = -1$ for each red point v_j, and we denote the m points by x_k, where $x_k = u_k$ for $k = 1, \ldots, p$ and $x_k = v_{k-p}$ for $k = p+1, \ldots, p+q$. Then the classification constraints can be written as

$$y_k(w^\top x_k - b) > 0 \quad \text{for } k = 1, \ldots, m.$$

The set of pairs $\{(x_1, y_1), \ldots, (x_m, y_m)\}$ is called a set of *training data* (or *training set*).

In the sequel, we will not use the above method, and we will stick to our two subsets of p *blue* points $\{u_i\}_{i=1}^p$ and q *red* points $\{v_j\}_{j=1}^q$.

Since there are infinitely many hyperplanes separating the two subsets (if indeed the two subsets are linearly separable), we would like to come up with a "good" criterion for choosing such a hyperplane.

The idea that was advocated by Vapnik (see Vapnik [182]) is to consider the distances $d(u_i, H)$ and $d(v_j, H)$ from *all* the points to the hyperplane H, and to pick a hyperplane H that maximizes the smallest of these distances. In machine learning this strategy is called finding a *maximal margin hyperplane*, or *hard margin support vector machine*, which definitely sounds more impressive.

Since the distance from a point x to the hyperplane H of equation $w^\top x - b = 0$ is

$$d(x, H) = \frac{|w^\top x - b|}{\|w\|},$$