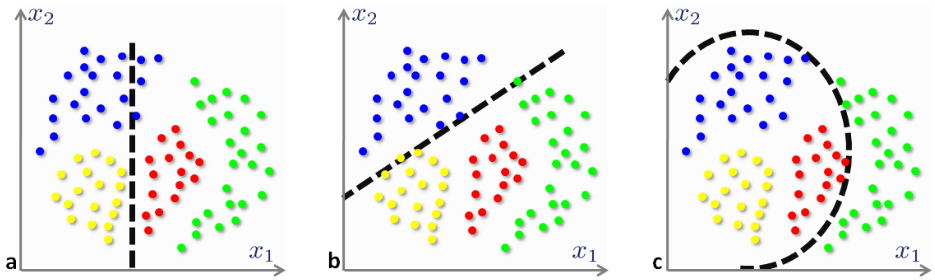


Fig. 2.6: **Example weak learners. (a)** Axis-aligned hyperplane. **(b)** General oriented hyperplane. **(c)** Quadratic (conic in 2D). For ease of visualization here we have $\mathbf{v} = (x_1\ x_2) \in \mathbb{R}^2$ and $\boldsymbol{\phi}(\mathbf{v}) = (x_1\ x_2\ 1)$ in homogeneous coordinates. In general data points $\mathbf{v}$ may have a much higher dimensionality and $\boldsymbol{\phi}$ still a dimensionality of $\leq 2$.

captures thresholds for the inequalities used in the binary test. The filter function $\boldsymbol{\phi}$ selects some features of choice out of the entire vector $\mathbf{v}$. All these parameters will be optimized at each split node. Figure 2.6 illustrates a few possible weak learner models, for example:

**Linear data separation.** In our model linear weak learners are defined as

$$h(\mathbf{v}, \boldsymbol{\theta}_j) = [\tau_1 > \boldsymbol{\phi}(\mathbf{v}) \cdot \boldsymbol{\psi} > \tau_2]\,, \tag{2.2}$$

where [.] is the indicator function[2]. For instance, in the 2D example in fig. 2.6b $\boldsymbol{\phi}(\mathbf{v}) = (x_1\ x_2\ 1)^\top$, and $\boldsymbol{\psi} \in \mathbb{R}^3$ denotes a generic line in homogeneous coordinates. In (2.2) setting $\tau_1 = \infty$ or $\tau_2 = -\infty$ corresponds to using a single-inequality splitting function. Another special case of this weak learner model is one where the line $\boldsymbol{\psi}$ is aligned with one of the axes of the feature space (*e.g.* $\boldsymbol{\psi} = (1\ 0\ \psi_3)$ or $\boldsymbol{\psi} = (0\ 1\ \psi_3)$, as in fig. 2.6a). Axis-aligned weak learners are often used in the boosting literature and they are referred to as *stumps* [98].

[2] Returns 1 if the argument is true and 0 if it is false.