Halfspaces

The **halfspace** hypothesis space is the set of hypotheses that consist of a hyperplane in a d-dimensional coordinate space that classifies a feature vector $\phi(x) \in \mathbb{R}^{d+1}$ as either -1 or 1 based on which side of the hyperplane it lies. Here d represents the number of features of item x. A hypothesis in this space is often called a **linear classifier** or a **perceptron**.

Mathematical Construction

A hyperplane in d-dimensional space is To rigorously define this hypothesis space, we first define the set of affine functions in d-dimensions as A_d where,

$$A_d = \{f_{\mathbf{w},b} \mid \mathbf{w} \in \mathbb{R}^d, b \in \mathbf{R}\}$$

and

$$f_{\mathbf{w},b}(\mathbf{v}) = \langle \mathbf{w}, \mathbf{v} \rangle + b$$

Since we are considering a coordinate vector space, we will use the dot product and thus,

$$f_{\mathbf{w},b}(\mathbf{v}) = \mathbf{w} \cdot \mathbf{v} + b$$

$$= \left(\sum_{i=1}^{d} w_i v_i \right) + b$$

An affine function is a hyperplane in d-dimensional space that is translated by a scalar b. In the context of affine functions for classification, the scalar b is called the **bias**.

Note that we can include the bias term into the inner-product by simply appending b to the beginning of the $\mathbf{w}$ vector to form $\mathbf{w}'$ and to then append the value 1 to the beginning of each $\mathbf{v}$ vector forming vector $\mathbf{v}$. That is, $v'_0 = 1$.

Then we see that,

$$\langle \mathbf{w}, \mathbf{v} \rangle + b = \langle \mathbf{w}', \mathbf{v}') \rangle$$

Proof:

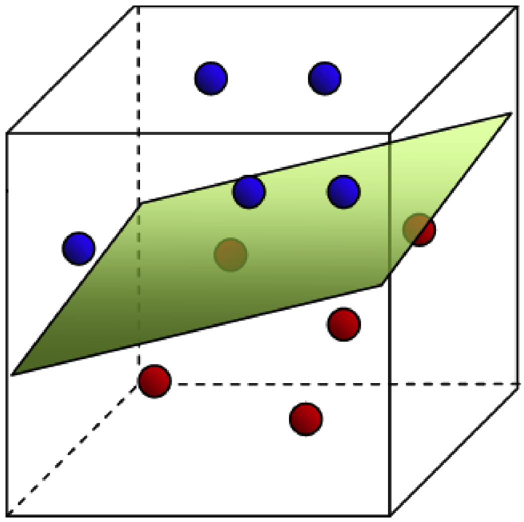

Figure 1: A hyperplane in $\mathbb{R}^3$ used as a decision-boundary for the instances that lie in this space.

$$
\begin{aligned}
f_{\mathbf{w},b}(\mathbf{v}) &= b + \langle \mathbf{w}, \mathbf{v} \rangle \\
&= b + \sum_{i=1}^{d} w_i v_i \\
&= b(1) + \sum_{i=1}^{d} w_i v_i \\
&= w_0^{'} v_0^{\prime} + \sum_{i=1}^{d} w_i^{\prime} v_i^{\prime} \\
&= \sum_{i=0}^{d} w_i^{\prime} v_i^{\prime} \\
&= \langle \mathbf{w}^{\prime}, \mathbf{v}^{\prime}) \rangle
\end{aligned}
$$

□

We will then redefine the notation so that the set of functions we consider will be linear functions rather than affine functions:

$$A_d = \{f_{\mathbf{w}} \mid \mathbf{w} \in \mathbb{R}^d\}$$

and

$$f_{\mathbf{w}}(\mathbf{v}) = \langle \mathbf{w}, \mathbf{v} \rangle$$

where $\mathbf{v}$ is an arbitrary vector in $\mathbb{R}^{d+1}$.

Classification

A half space hypothesis space $\mathcal{H}$ is then the set of functions

$$\mathcal{H} = \{\text{sign}(f_{\mathbf{w}}(\mathbf{v})) \mid f_{\mathbf{w}} \in A_d\}$$

That is, each hypothesis in $\mathcal{H}$ is parameterized by coefficients $w_0, w_1, \ldots, w_d$ that defines the hyperplane decision boundary between -1 and 1 instances.

Also, we note that the weight vector $\mathbf{w}$ is normal to the hyperplane. If we have a vector $\mathbf{v}$ that lies on the hyperplane, then their inner product is zero. That is,

$$\langle \mathbf{w}, \mathbf{v} \rangle = 0 \implies \mathbf{w} \perp \mathbf{v}$$

When operating on feature vectors $\phi(x) \in \mathbb{R}^{d+1}$, we will always assume that $\phi(x)_0 = 1$ and that $\phi(1)_1 \ldots \phi(x)_d$ are the d features of item x. Furthermore we will assume that w_0 of $\mathbf{w}$ will be the bias term.

Lastly, we see that instances that fall on the same side of the hyperplane for which $\mathbf{w}$ points will form an acute angle from the hyperplane to $\mathbf{w}$. Conversely, instances that fall on the opposite side of the hyperplane for which $\mathbf{w}$ points will form an obtuse angle from the hyperplane to $\mathbf{w}$. Thus, the side of the hyperplane for which $\mathbf{w}$ points away from will be the side of the hyperplane with positive labeled instances.