The Margin of a Separating Hyperplane

We are given the following setting:

- An inner product space $\mathcal{V}$ with norm $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$
- Labelled vectors $(y_1, \mathbf{x}_1), \ldots, (y_m, \mathbf{x}_1)$
- A separating hyperplane define by the vector $\mathbf{w}$ where $\mathbf{w}$ is normal to the hyperplane with norm $||\mathbf{w}|| = 1$. This hyperplane separates all of the labelled vectors by their label. That is
$$\forall i \; y_i \langle \mathbf{w}, \mathbf{x}_i \rangle > 0$$

In this setting, the **margin** of the hyperplane is defined to be the closest vector to the separating hyperplane. This margin is given as follows:

$$margin = \min_i \{ |\langle \mathbf{w}, \mathbf{x}_i \rangle| \}$$

As we will soon prove, $|\langle \mathbf{w}, \mathbf{x}_i \rangle|$ is the distance of $\mathbf{x}_i$ to the hyperplane. Thus, we see that the margin is defined to be the minimal distance of any point to the hyperplane.

Distance of a Vector to the Hyperplane

The distance of a point $\mathbf{x}$ to the hyperplane is defined as

$$\min \{ ||\mathbf{x} - \mathbf{v}|| \mid \langle \mathbf{w}, \mathbf{v} \rangle = 0 \}$$

That is, the distance to hyperplane is the smallest distance between $\mathbf{x}$ and some vector $\mathbf{v}$ that lies on the hyperplane (i.e. $\mathbf{v}$ lies on the hyperplane because it is orthogonal to $\mathbf{w}$ as evident by the fact $\langle \mathbf{w}, \mathbf{v} \rangle = 0$). In fact, this distance is given by

$$|\langle \mathbf{w}, \mathbf{x} \rangle + b| = \min \{ ||\mathbf{x} - \mathbf{v}|| \mid \langle \mathbf{w}, \mathbf{v} \rangle = 0 \}$$

Proof:

Let the vector $\mathbf{v}$ be defined to be

$$\mathbf{v} := \mathbf{x} - \langle \mathbf{w}, \mathbf{x} \rangle \mathbf{w}$$

We see that this vector lies on the hyperplane because it's inner product with $\mathbf{w}$ is 0 as shown here:

$$\begin{aligned}\langle \mathbf{w}, \mathbf{v}\rangle &= \langle \mathbf{w}, \mathbf{x} - \langle \mathbf{w}, \mathbf{x}\rangle \mathbf{w}\rangle \\ &= \langle \mathbf{w}, \mathbf{x}\rangle - \langle \mathbf{w}, \langle \mathbf{w}, \mathbf{x}\rangle \mathbf{w}\rangle \\ &= \langle \mathbf{w}, \mathbf{x}\rangle - \langle \mathbf{w}, \mathbf{x}\rangle\langle \mathbf{w}, \mathbf{w}\rangle \\ &= \langle \mathbf{w}, \mathbf{x}\rangle - \langle \mathbf{w}, \mathbf{x}\rangle \, ||\mathbf{w}||^2 \\ &= \langle \mathbf{w}, \mathbf{x}\rangle - \langle \mathbf{w}, \mathbf{x}\rangle \\ &= 0\end{aligned}$$

Next we see that no vector on the hyperplane is closer to $\mathbf{x}$ as $\mathbf{v}$.

$$\begin{aligned}||\mathbf{x} - \mathbf{u}||^2 &= ||\mathbf{x} - \mathbf{v} + \mathbf{v} - \mathbf{u}||^2 \\ &= ||\mathbf{x} - \mathbf{v}||^2 + ||\mathbf{v} - \mathbf{u}||^2 + 2\langle \mathbf{x} - \mathbf{v}, \mathbf{v} - \mathbf{u}\rangle \quad \text{See Lemma 1.} \\ &\geq ||\mathbf{x} - \mathbf{v}||^2 + 2\langle \mathbf{x} - \mathbf{v}, \mathbf{v} - \mathbf{u}\rangle \\ &= ||\mathbf{x} - \mathbf{v}||^2 + 2\langle \mathbf{x} - (\mathbf{x} - \langle \mathbf{w}, \mathbf{x}\rangle \mathbf{w}), \mathbf{v} - \mathbf{u}\rangle \\ &= ||\mathbf{x} - \mathbf{v}||^2 + 2\langle \mathbf{x} - \mathbf{x} + \langle \mathbf{w}, \mathbf{x}\rangle \mathbf{w}, \mathbf{v} - \mathbf{u}\rangle \\ &= ||\mathbf{x} - \mathbf{v}||^2 + 2\langle\langle \mathbf{w}, \mathbf{x}\rangle \mathbf{w}, \mathbf{v} - \mathbf{u}\rangle \\ &= ||\mathbf{x} - \mathbf{v}||^2 + 2\langle \mathbf{w}, \mathbf{x}\rangle\langle \mathbf{w}, \mathbf{v} - \mathbf{u}\rangle \\ &= ||\mathbf{x} - \mathbf{v}||^2 + 2\langle \mathbf{w}, \mathbf{x}\rangle\langle \mathbf{w}, \mathbf{v} - \mathbf{u}\rangle \\ &= ||\mathbf{x} - \mathbf{v}||^2 \quad \mathbf{v} - \mathbf{u} \text{ is on the hyperplane, thus } \langle \mathbf{w}, \mathbf{v} - \mathbf{u}\rangle = 0\end{aligned}$$

Lastly, we see that the norm of $\mathbf{x} - \mathbf{v}$ is $\langle \mathbf{w}, \mathbf{x}\rangle$:

$$\begin{aligned}||\mathbf{x} - \mathbf{v}|| &= ||\mathbf{x} - (\mathbf{x} - \langle \mathbf{w}, \mathbf{x} \rangle \mathbf{w})|| \\ &= ||\langle \mathbf{w}, \mathbf{x} \rangle \mathbf{w})|| \\ &= \langle \mathbf{w}, \mathbf{x} \rangle \, ||\mathbf{w})|| \\ &= \langle \mathbf{w}, \mathbf{x} \rangle\end{aligned}$$

□

Lemma 1:

$$||\mathbf{v} + \mathbf{u}||^2 = ||\mathbf{v}||^2 + ||\mathbf{u}||^2 + 2\langle \mathbf{v}, \mathbf{u} \rangle$$

Proof of Lemma 1:

$$\begin{aligned}||\mathbf{v} + \mathbf{u}||^2 &= \langle \mathbf{v} + \mathbf{u}, \mathbf{v} + \mathbf{u} \rangle \\ &= \langle \mathbf{v}, \mathbf{v} + \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{v} + \mathbf{u} \rangle \\ &= \langle \mathbf{v}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{v} + \mathbf{u} \rangle \\ &= \langle \mathbf{v}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{u} \rangle \\ &= ||\mathbf{v}||^2 + \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle + ||\mathbf{u}||^2 \\ &= ||\mathbf{v}||^2 + ||\mathbf{u}||^2 + 2\langle \mathbf{v}, \mathbf{u} \rangle\end{aligned}$$

◇