# Computing the margin width

Francesco Faccio
francesco@idsia.ch

Let your model be $f(x) = w^T \phi(x) + b$, where x is an input vector, w and b are weights and bias, $\phi(\cdot)$ is a fixed feature transformation.

Consider two points $\phi(x_1), \phi(x_2)$ lying on the plus plane:

$$w^T \phi(x_1) + b = 1 \tag{1}$$

$$w^T \phi(x_2) + b = 1 \tag{2}$$

which implies:

$$w^T (\phi(x_1) - \phi(x_2)) = 0 \tag{3}$$

i.e. the vector $w$ is orthogonal to the decision boundary.

Consider now a point in the minus plane $\phi(x^-)$ and the closest point in the plus plane $\phi(x^+)$. Since w is parallel to the vector connecting $\phi(x^-)$ and $\phi(x^+)$, then it holds that

$$\phi(x^+) = \phi(x^-) + \lambda w \tag{4}$$

for some $\lambda$.

Furthermore, for points $\phi(x^-)$ and $\phi(x^+)$ we have:

$$w^T \phi(x^+) + b = 1 \tag{5}$$

$$w^T \phi(x^-) + b = -1 \tag{6}$$

And the margin M is defined as the distance between these two points

$$M = ||\phi(x^+) - \phi(x^-)|| \overset{(4)}{=} ||\lambda w|| \tag{7}$$

Substituting (4) into (5) we have

$$w^T (\phi(x^-) + \lambda w) + b = 1 \Leftrightarrow \tag{8}$$

$$w^T \phi(x^-) + w^T \lambda w + b = 1 \overset{(6)}{\Leftrightarrow} \tag{9}$$

$$-1 - b + w^T \lambda w + b = 1 \Leftrightarrow \lambda w^T w = 2 \Leftrightarrow \lambda = \frac{2}{w^T w} \tag{10}$$

Finally:

$$M \overset{(7)}{=} ||\lambda w|| = \lambda ||w|| \overset{(10)}{=} \frac{2\sqrt{w^T w}}{w^T w} = \frac{2}{\sqrt{w^T w}} = \frac{2}{||w||} \tag{11}$$