Support Vector Machines

Introduction

Support vector machines are a widely used supervised learning model using for classification and regression. SVM attempts to fit a hyperplane which maximizes the distance separating samples from the decision boundary. By exploiting kernels, SVM can perform non-linear classification by implicitly mapping inputs into higher dimensions.

Margins

Binary classification

We want to learn a function that maps $x \in \mathbb{R}^d \to \{-1, -1\}$. Denote our dataset as $\mathcal{D} = \{x_i, y_i\}$, for $i \in [1, \ldots, n]$.

Let's derive the equation of a plane. Assume we have some plane in some vector space. A point $P = \langle x, y, z \rangle$ lies on that plane. The normal to the plane is $\vec{w}$. Another point $P_0 = \langle x_0, y_0, z_0 \rangle$ lies on the plane. The vector from P to P_0 is given by $\vec{P_0} - \vec{P}$ is equal to $\langle x_0 - x, y_0 - y, z_0 - z \rangle$. Since this is orthogonal to w (by definition):

$$
\begin{aligned}
\vec{w} \cdot (\vec{P_0} - \vec{P}) &= 0 \\
\vec{w} \cdot \vec{P_0} - \vec{w} \cdot \vec{P} &= 0 \\
\vec{w} \cdot \vec{P} + b &= 0
\end{aligned}
$$

Consider the vector space of our data points, x_i for $i \in [1, n]$. We describe a separating hyperplane as $w^T x + w_0 = 0$. Here, w is the normal vector of the plane, and w_0 is the Euclidean distance from the origin to the plane.

The distance from a point x_i to the hyperplane is given by:

$$
\frac{|w^T x_i + w_0|}{||w||} = \frac{1}{||w||}
$$