Constructing a Support Vector Machine

Jonathan Enders

November 26, 2015

Abstract

Support vector machines can be implemented as supervised machine learning algorithms in which differently labeled data is used to position planes in the training data feature space. These planes can then be used to classify further data giving it suitable labels.

The plane that separates the two differently classified regions should be parameterized in terms of only one vector $\vec{n}$ that points from the origin to a point on the plane. Every possible plane[1] can be represented this way if the convention is used that the plane stands perpendicular on that vector.
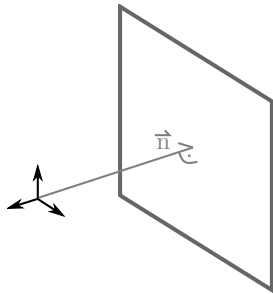

Figure 1: A normal vector determines the configuration of the plane

The coordinates of a point from the plane p_l can be calculated using the coordinates of the point from the origin $\vec{p}$.

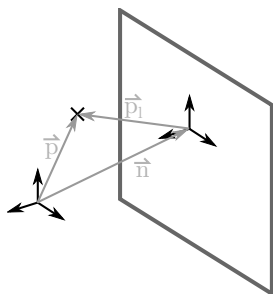$$\vec{p_l} = \vec{p} - \vec{n}$$

Figure 2: Coordinates of the point

To calculate the shortest distance d of a point from the plane equation (1) can be used.

$$d = \frac{\vec{n} \cdot \vec{p_l}}{|\vec{n}|} \tag{1}$$

---

[1]Planes going through the origin are a problem in this description because the zero vector doesn't yield any directional information, i.e. the direction perpendicular to the zero vector is not defined.
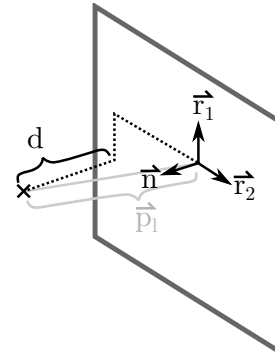

Figure 3: How to obtain the distance between a point and a plane

Thus the distance of all the points from the plane could be calculated as follows

$$
\begin{aligned}
d_{\text{total}} &= \sum_p \frac{\vec{n} \cdot (\vec{p} - \vec{n})}{|\vec{n}|} \\
&= \left(\sum_p \frac{\vec{n} \cdot \vec{p}}{|\vec{n}|} \right) - |\vec{n}| * N
\end{aligned}
$$

with N being the amount of points.

The plane is to be positioned in such a way that it separates the differently labeled data points. At the same time it ought to be as centered between the classes of points as possible, i.e. the distance from the points to the plane has to be maximal still fulfilling the upper condition. Minimizing the following function (2) with respect to the coordinates of the plane yields a plane that fulfills the upper conditions.

$$\sum_p \exp \left(c_p \left(\frac{\vec{n} \cdot \vec{p}}{|\vec{n}|} - |\vec{n}| \right) \right) \tag{2}$$

$c_p \in \{-1, +1\}$ is a factor that is positive for one class of points and negative for the other class as determined by the different labeling of the training data set.