Support Vector Machine

1 Linear Separable SVM

Definition: Hyperplane

A subspace whose dimension is one less than its ambient space

For the context of this lecture, the ambient space is defined as the Hilbert space **Intuition**
Given training data (x_i, y_i) for $i = 1, ..., n$ with $x_i \in \mathbb{R}^2$ and $y_i \in \{-1, +1\}$ learn a classifier $f(x)$ training such that

$$f(x_i) = \begin{cases} \geq 0, & y_i = +1 \\ < 0, & y_i = -1 \end{cases}$$

The problem with this solution is that there is no optimal solution of the Hyperplane given the training data points. Hyperplane have can alternately be named decision boundary. Many lines could be just as valid

Definition: Separating Hyperplane

Let $S = \{(x_i, y_i)\}_{i=1}^m \in \mathbb{R}^d \times \{-1, +1\}$ be a training set
By a hyperplane we mean a set of Hilbert space $H_{w,b} = \{x \in \mathbb{R}^d : w^T x + b = 0\}$ parametrised by $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$
We assume that the data are linearly separable, that is, there exist $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$ such that
$y_i(w^T x_i + b) > 0, i = 1, .., m$
In which case we call $H_{w,b}$ a separating hyperplane

Definition: Distance

The distance $\rho_x(w, b)$ of a point x from a hyperplane $H_{w,b}$ is

$$\rho_x(w, b) = \frac{\left|w^T x + b\right|}{||w||}$$

Definition: Margin

If $H_{w,b}$ separates the training set S we define its margin as:

$$\rho_x(w, b) = \min_{i=1:m} \rho_{x_i}(w, b)$$

If $H_{w,b}$ is a hyperplane (separating or not) we also define the margin of a point x as $w^T x + b$

1.1 Optimal separating hyperplane

The separating hyperplane with maximum margin can be solved with the following optimisation problem

$$\rho(S) = \max_{w,b} \min_i \left\{ \frac{y_i\left(w^T x_i + b\right)}{||w||} : y_i\left(w^T x_i + b\right) \geq 0, \quad i = 1, \ldots, m \right\} > 0$$

A separating hyperplane is parameterised by (w,b), but this choice is not unique (rescaling with a positive constant gives the same separating hyperplane)

There are two possible ways to fix the parameterisation:

- **Normalised hyperplane**: set $||w|| = 1$, in which case $\rho_x(w, b) = |w^T x + b|$ and $\rho_s(w, b) = \min_{i=1:m} y_i(w^T x_i + b)$

- **Canonical hyperplane**: choose $||w||$ such that $\rho_s(w,b) = \frac{1}{||w||}$, i.e. we require that $\min_{i=1:m} y_i(w^T x_i + b) = 1$

The problem thus can be defined as

Minimise $\frac{1}{2}w^T w$

Subject to $y_i(W^T x_i + b) \geqslant 1, i = 1, ..., m$

1.2 Saddle point

Definition: Saddle point

A point on the surface og a graph of a function where the slopes in orthogonal directions are all zero, but which is not a local extremum of the function

To determine the saddle point of the Lagrangain function

$$L(w,b;\alpha) = \frac{1}{2}w^T w - \sum_{i=1}^{m} \alpha_i \left\{ y_i \left(w^T x_i + b \right) - 1 \right\}$$

where $\alpha_i \geqslant 0$ are the Lagrange multipliers

We minimise L over (w,b) and maximise over α. Differentiating with respect to w and b we obtain

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{m} y_i \alpha_i = 0$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{m} \alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^{m} \alpha_i y_i x_i$$