Perceptron

NOTE: Github is not support latex well , so I upload Readme in PDF.

Two classes classify

The simplest representation of a linear discriminant function is obtained by taking a linear function of the input vector so that

$$y(\mathbf{x}) = \mathbf{w}^\intercal \mathbf{x} + w_0 \tag{1}$$

where $\mathbf{w}$ is called a **weight vector**,and w_0 is a bias. The negative of the bias is sometimes called a threshold. An input vector $\mathbf{x}$ assign to class C_1 if $y(\mathbf{x}) \geq 0$ and to class C_2 otherwise. The corresponding decision boundary is therefore defined by the relation $y(\mathbf{x}) = 0$.which corresponds to a **(D-1)-dimensional hyperplane within D-dimensional input space**. the vector $\mathbf{w}$ is ortheogonal to every vector lying within the decision surface, and so the normal distance from the origin to the decision surface is given by

$$\frac{\mathbf{w}^\intercal \mathbf{x}}{\|\mathbf{w}\|} = -\frac{w_0}{\|\mathbf{w}\|} \tag{2}$$

We therefore see that the bias parameter w_0 determines the location of the decision surface.

Perceptron

- **Introduction**

  Perceptron corresponds to a two-class model in which the input vector $\mathbf{x}$ is first transformed using a fixed nonlinear transformation to give a feature vector $\phi(\mathbf{x})$, and this is then used to construct a generalized linear model of the form where the nolinear activation function $f(\cdot)$ is given by a step function of the form

$$y(x) = f(\mathbf{w}^\intercal \phi(\mathbf{x})) \tag{3}$$

Where nolinear activation function $f(\cdot)$ is given by a step function of the form

$$f(a) = \begin{cases} +1, & (a \geq 0) \\ -1, & (a \leq 0) \end{cases} \tag{4}$$

In early discussions of two-class classification problems, we have focused on a target coding scheme in

wich $t \in 0, 1$, which is appropriate in the context of probabilistic models. For perceptron, however, it is

more convinient to use target values $t = +1$ for class C_1 and $t = -1$ for class C_2, which matches the choice of activation function.

- **Error Measure**

According to equation (4), it is easy to find that $\mathbf{w}^\intercal \phi(x_n) t_n > 0$ if x_n was classified correctly. The perceptron criterion is therefore given by

$$E_p(\mathbf{w}) = -\sum_{n \in M} \mathbf{w}^\intercal \phi_n t_n \tag{5}$$

Where the M denotes the set of all misclassified patterns.

so we now apply the stochastic gradient sescent algorithm to this error function. The change in the weight vector $\mathbf{w}$ is the given by

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta \nabla E_p(\mathbf{w}) = \mathbf{w}^{(\tau)} + \eta \phi_n t_n \tag{6}$$

- **Demonstration of convergence**
    - TODO