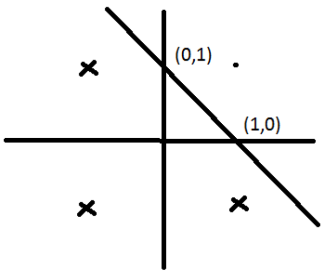
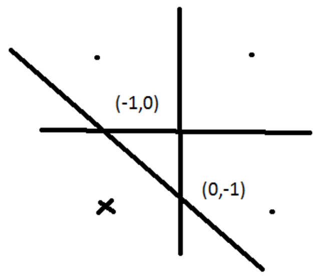
Machine Learning Homework 3-Saket Vishwasrao

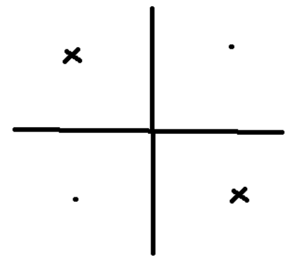
Written Problems



AND



OR



XOR

No linear classifier can possibly exist

1. (a) AND can be expressed as a form of linear classifier with $w_1 = 1, w_2 = 1, b = -1$
   Or can also be expressed as a linear classfier with $w_1 = 1, w_2 = 1, b = 1$
   XOR cannot be expressed as a linear classifier.

   (b) For expressing XOR as a 2 layered perceptron,we can split it as $h_{11} = AND(\bar{x_1}, x_2), h_{12} = AND(x_1, \bar{x_2})$ The second layer is simply the $OR(h_{11}, h_{12})$
   where $\mathbf{h_1} = [h_{11}, h_{12}]$

   Thus $\mathbf{w^{(1)}} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ and $\mathbf{b^{(1)}} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

   $\mathbf{w^{(2)}} = \begin{pmatrix} 1 & 1 \end{pmatrix}$ and $\mathbf{b^{(2)}} = (1)$

2. (a) For the points $\mathbf{x_-}$ and $\mathbf{x_+}$ to lie on the margin, we have

$$\mathbf{w^T x_-} = -1, \mathbf{w^T x_+} = 1, \mathbf{x_+} = \mathbf{x_-} + \gamma \mathbf{w} \tag{1}$$

   Combining the above equations,

$$\mathbf{w^T}(\mathbf{x_-} + \gamma \mathbf{w}) = 1 \tag{2}$$

$$\implies \mathbf{w^T x_-} + \mathbf{w^T} \gamma \mathbf{w} = 1 \tag{3}$$

$$\implies \mathbf{-1} + \gamma \mathbf{w^T w} = \mathbf{1} \tag{4}$$

$$\implies \gamma = \frac{2}{\mathbf{w^T w}} \tag{5}$$

$$margin = \sqrt{(\mathbf{x_+} - \mathbf{x_-})^\mathbf{T}(\mathbf{x_+} - \mathbf{x_-})} \tag{6}$$

$$= \sqrt{\left(\frac{\mathbf{2w}}{\mathbf{w^Tw}}\right)^\mathbf{T}\left(\frac{\mathbf{2w}}{\mathbf{w^Tw}}\right)} \tag{7}$$

$$= \sqrt{\frac{\mathbf{4w^Tw}}{\mathbf{w^Tw^2}}} \tag{8}$$

$$= \frac{2}{\sqrt{\mathbf{w^Tw}}} \tag{9}$$

(b) Thus the objective is

$$\arg\max_{w} \frac{2}{\sqrt{\mathbf{w^Tw}}} \tag{10}$$

$$\text{such that } y_i(\mathbf{w^Tx_i} - 1) \geqslant 0$$

This is equivalent to

$$\arg\min_{w} \frac{1}{2}\mathbf{w^Tw} \tag{11}$$

$$\text{such that } y_i(\mathbf{w^Tx_i} - 1) \geqslant 0$$

Adding slack penalty

$$\arg\min_{w,\xi\in[0,\infty)} \frac{1}{2}\mathbf{w^Tw} + C\sum_{i=1}^{n}\xi_i \tag{12}$$

$$\text{such that } y_i(\mathbf{w^Tx_i}) \geqslant 1 - \xi_i$$

Writing this optimisation equation in form of a Langrange Multiplier

$$\arg\min_{w\in\mathbb{R}^d,\xi\in\mathbb{R}^n_+} \arg\max_{\alpha\in[0,\infty]^n,\beta\in[0,\infty]^n} L(w,\xi,\alpha,\beta) \tag{13}$$

$$L(w,\xi,\alpha,\beta) = \frac{1}{2}\mathbf{w^Tw} + C\sum_{i=1}^{n}\xi - \sum_i \alpha_i(y_i(\mathbf{w^Tx_i}) - 1 + \xi_i) - \sum_i \beta_i\xi_i \tag{14}$$

(c) The dual objective of the Langrangian becomes

$$\arg\max_{\alpha\in[0,\infty]^n,\beta\in[0,\infty]^n} \arg\min_{w\in\mathbb{R}^d,\xi\in\mathbb{R}^n_+} L(w,\xi,\alpha,\beta) \tag{15}$$

The gradient of the Langrangian w.r.t to w is

$$\nabla_w L = \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x_i} \tag{16}$$

$$\implies \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x_i} = 0 \tag{17}$$

as at the optimum the gradient is zero.Thus

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x_i} \tag{18}$$

(d)

$$L(w,\xi,\alpha,\beta) = \frac{1}{2}\mathbf{w^T w} + C\sum_{i=1}^{n}\xi_i - \sum_i \alpha_i(y_i(\mathbf{w^T x_i}) - 1 + \xi_i) - \sum_i \beta_i\xi_i \tag{19}$$

$$L(w,\xi,\alpha,\beta) = \frac{1}{2}\mathbf{w^T w} + C\sum_{i=1}^{n}\xi_i - \sum_i \alpha_i y_i \mathbf{w^T x_i} + \sum_i \alpha_i - \sum_i \alpha_i\xi_i - \sum_i \beta_i\xi_i \tag{20}$$

Taking the gradient w.r.t to ξ

$$\nabla_\xi L = C - \alpha - \beta = 0 \tag{21}$$

Substituting the value of w from equation 18

$$L(w,\xi,\alpha,\beta) = \frac{1}{2}\mathbf{w^T}\sum_i \alpha_i y_i \mathbf{x_i} + C\sum_{i=1}^{n}\xi_i - w^T\sum_i \alpha_i y_i \mathbf{x_i} + \sum_i \alpha_i - \sum_i \alpha_i\xi_i - \sum_i \beta_i\xi_i \tag{22}$$

$$= -\frac{1}{2}\mathbf{w^T}\sum_i \alpha_i y_i \mathbf{x_i} + C\sum_{i=1}^{n}\xi_i + \sum_i \alpha_i - \sum_i \alpha_i\xi_i - \sum_i \beta_i\xi_i \tag{23}$$

$$= -\frac{1}{2}\mathbf{w^T}\sum_i \alpha_i y_i \mathbf{x_i} + \sum_i \alpha_i - \sum_i (C-\alpha_i)\xi_i - \sum_i \beta_i\xi_i \tag{24}$$

$$= -\frac{1}{2}\mathbf{w^T}\sum_i \alpha_i y_i \mathbf{x_i} + \sum_i \alpha_i \tag{25}$$

since $C - \alpha_i = \beta_i$ from equation 21

$$= -\frac{1}{2}\left(\sum_i \alpha_i y_i \mathbf{x_i}\right)^T \sum_j \alpha_j y_j \mathbf{x_j} + \sum_i \alpha_i \tag{26}$$

$$= -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \mathbf{x_i^T x_j} + \sum_i \alpha_i \tag{27}$$

now from equation 21,we have $\alpha = C - \beta$ and $\beta = C - \alpha$ and since α, β is positive we have that $\alpha \in [0, C]$
Thus the langrangian becomes

$$\arg\max_\alpha -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \mathbf{x_i^T x_j} + \sum_i \alpha_i \tag{28}$$

such that $\alpha_i \in [0, C]$

This is equivalent to

$$\arg\min_\alpha \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \mathbf{x_i^T x_j} - \sum_i \alpha_i \tag{29}$$

such that $\alpha_i \in [0, C]$

and

$$f(x) = \mathbf{w^T x} = \sum_i \alpha_i y_i \mathbf{x_i^T x} \tag{30}$$

(e) Using $K(x_i, x_j) = \mathbf{x_i^T x_j}$ the langrangian becomes

$$\arg\min_{\alpha} \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_i \alpha_i \tag{31}$$

$$\text{such that } \alpha_i \in [0, C]$$

and

$$f(x) = \mathbf{w^T x} = \sum_i \alpha_i y_i K(x_i, x) \tag{32}$$