COMPUTER SCIENCE TRIPOS Part IB – 2024 – Paper 7

1 Artificial Intelligence (sbh11)

For a two-class classification problem with classes C_1 and C_2 , we will use the following linear classifier with probabilistic output

$$\Pr(C_1|\mathbf{x}, \mathbf{w}, w_0, \theta) = \sigma_\theta \left(\mathbf{w}^T \mathbf{x} + w_0 \right)$$
(1)

where **x** is an input vector and **w**, w_0 and θ are parameters, and

$$\sigma_{\theta}(x) = \frac{1}{1 + \exp(-\theta x)}$$

is the activation function.

- (a) The activation function σ_{θ} often appears assuming $\theta = 1$. What role does θ play when its value can be set freely? [1 mark]
- (b) You are provided with a collection $\mathbf{s} = ((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m))$ of m training examples where $y_i \in \{0, 1\}$ for $i = 1, \dots, m$. You suspect that the data has a band-like structure. In two dimensions:



Explain how the linear classifier can be modified, changing only its activation function and without adding further parameters, such that it can be trained on this type of data. It should be possible to set the width of the C_1 region as part of the training. [6 marks]

(c) When using a two-class classifier $h(\mathbf{x}, \mathbf{p})$ that has parameters \mathbf{p} and outputs probabilities, one often uses the error measure

$$E(\mathbf{p}) = -\sum_{i=1}^{m} y_i \log h(\mathbf{x}_i, \mathbf{p}) + (1 - y_i) \log(1 - h(\mathbf{x}_i, \mathbf{p})).$$

Explain how the gradient descent method can be applied to solve this classification problem for the data described in Part (b), giving explicit formulae for the gradients. [13 marks]