CAS CS 640 Homework 1

Alec Hoyland (U83403624)

October 9, 2019

Question 1.

(a) accuracy = (TP + TN)/(P + N) = (16 + 8)/(20 + 10) = 24/30 = 0.8.

(b) The point in ROC space is at $(FP/(FP+TN), TP/(TP+TN))$. Therefore, $(4/(4+8), 16/(16+8)) = (0.33, 0.67)$.

(c) The F-1 score is $2TP/(2TP+FP+FN)$. Therefore, $F_1 = 2*16/(2*16+4+2) = 0.84$

Question 2.

In this neural network, the activation function is an x-shifted Heaviside function.

(a) The input vector is $[0, 0, 1, 0, 0, 1]$. This leads to a hidden layer state of $[2, 2]$, which is activated to $[1, 1]$. This leads to an output layer state of $[4, -4]$, which is activated to $[1, 0]$. Thus, the neural network concludes that Romeo and Juliet are acquaintances.

(b) All the weights are going to be multiplied by a constant C. Since the activation functions are linear and the other operations are inner products (which are linear), the neural network mapping from input to output is itself a linear mapping. Since homogeneity is a property of all linear systems, for any linear mapping f, $f(Cx) = Cf(x)$. This tells us that the output vector as a function of the input vector is $\vec{y} = f(\vec{x})$, which tells us that for $f(C\vec{x})$, the resultant vector $C\vec{y}$, when normalized, results in the same probabalistic interpretation as the unscaled neural network.

Here is an example, with the same input vector as before.

The input vector is $[0, 0, 1, 0, 0, 1]$. This leads to a hidden layer state of $[2C, 2C]$, which is activated to $[1, 1]$. This leads to an output layer state of $[4C, -4C]$, which is activated to $[1, 0]$. Thus, the neural network concludes that Romeo and Juliet are acquaintances.

Question 3.

(a)

$$\frac{\partial P}{\partial o_z} = \frac{1 - d_z}{1 - o_z} - \frac{d_z}{o_z}$$

(b) Since $o_z(w) = \sigma(w \cdot o_h)$ (I am omitting any indices here, so $\cdot$ represents scalar multiplication),

$$\frac{\partial o_z}{\partial w} = \sigma'(w \cdot o_h)o_h = \sigma(w \cdot o_h)(1 - \sigma(w \cdot o_h))o_h$$

(c)

$$\frac{\partial P}{\partial w} = \frac{\partial P}{\partial o_z}\frac{\partial o_z}{\partial w} = \left(\frac{1 - d_z}{1 - o_z} - \frac{d_z}{o_z}\right)\Big(\sigma(w \cdot o_h)(1 - \sigma(w \cdot o_h))o_h\Big)$$

(d)

$$w_{new} = w - r \cdot \frac{\partial P}{\partial w}$$

(e)

$$w_{new} = 0.4 - 0.1 \cdot \left(\frac{1 - 0.9}{1 - 0.7} - \frac{0.9}{0.7}\right)\Big(\sigma(0.4 \cdot 0.8)\Big(1 - \sigma(0.4 \cdot 0.8)\Big)0.8\Big)$$

$$w_{new} = 0.4 - 0.1 \cdot \left(\frac{0.1}{0.3} - 1.3\right)\Big(0.58(1 - 0.58)0.8\Big)$$

$$w_{new} = 0.4 - 0.1 \cdot -0.97 \cdot 0.195$$

$$w_{new} = 0.42$$

Colophon

This document was typeset using the $\LaTeX\,2_\varepsilon$ document processing system originally developed by Leslie Lamport, based on the $\TeX$ typesetting system created by Donald Knuth. The class is `latex-homework-class` by Jake Zimmerman, released under the MIT license. All above work was done by the authors. I used Bishop's machine learning textbook as reference.