CSC 422 HW3

Vinhson Phan, Sagnik Nayak, Rhea John

April 2021

1. (a) $P(A) = P(C) * P(A|C) + P(\sim C) * P(A| \sim C) = .6 * .4 + .4 * .5 = .44$
   (b) $P(D|B, \sim A) = P(D|B) * P(\sim A) = .3 * .56 = .168$
   (c) $P(A, B, \sim C, D, E, F) = P(\sim C) * P(A| \sim C) * P(F| \sim C) * P(B) * P(D|B) * P(E| \sim C, D) = .4 * .5 * .2 * .4 * .3 * .1 = .00048$
   (d) We have that E and F are conditionally independent since we do know C which is a parent of both nodes and then using the Markov condition they E and F are nondescendants and thus conditionally independent.
   (e) Yes, because $P(A|B) = P(A)$ and $P(B|A) = P(B)$ or in other words knowing 1 of the probabilities does not affect the other.
   (f) E is conditionally independent of both F, A given C, D. So $P(E) = P(E|C, \sim D) = .4$ and thus $P(\sim E|C, \sim D) = .6$. It is more likely for E to be false.

2. (a) We can find the weights of $w_0 + w_1\sqrt{x}$ by computing the equation $w = (X^T X)^{-1} X^T y$ where

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 3 \end{bmatrix}, y = \begin{bmatrix} 5 \\ -2 \\ -3 \\ -4 \end{bmatrix}$$

   The first column of X corresponds to a coefficient of 1 to each w_0 and the second column is the square root of each x_i in the dataset.

   Furthermore we compute $X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}$

   Then $(X^T X)^{-1} = \begin{bmatrix} \frac{7}{10} & -\frac{3}{10} \\ -\frac{3}{10} & \frac{1}{5} \end{bmatrix}$ which can be computed by taking the adjoint over the determinant. Then we can commpute through matrix multiplication that

$$\begin{bmatrix} \frac{7}{10} & -\frac{3}{10} \\ -\frac{3}{10} & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ -2 \\ -3 \\ -4 \end{bmatrix} = \begin{bmatrix} -1.3 \\ .2 \end{bmatrix} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

   (b) Our linear regression model is now given by $-1.3 + .2\sqrt{x}$. Plugging in for each predicted y-value, $\hat{y}$, we get $\hat{y_1} = -.9$, $\hat{y_2} = -1.3$, $\hat{y_3} = -1.1$, $\hat{y_4} = -.7$
   Then we compute RMSE as follows

$$\text{RMSE} = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n}} = \sqrt{\frac{5.9^2 + (-.7)^2 + (-1.9)^2 + (-3.3)^2}{4}} = 3.53$$

3. (a) Using forward pass, we can compute the activation of H1 and H2 neurons using the equations H1 = X1(w1) + X2(w3) and H2 = X1(w2) + X2(w4). This results in H1 = -0.5 and H2 = -0.2. Using the new found values of H1 and H2, we can use the following equation to compute the activation of O: H1(w5) + H2(w6). The calculated activation of O is -0.13.

(b) Backward Pass

i. Equations

A. Cost Equation:

$$\mathrm{C} = (Y_i - a_o)^2$$

B. Activation of final layer:

$$a_o = a_{H1}(W_5) + a_{H2}(W_6)$$

C. Activation of node aH1:

$$a_{H1} = W_1(X_1) + W_3(X_2)$$

ii.

$$\frac{\partial C}{\partial a_o} = (Y_i - a_o) = -2(1 - (-0.13)) = -2.26$$

$$\frac{\partial a_o}{\partial W_5} = Mz = 1(a_{H1}) = -0.5$$

$$\frac{\partial a_o}{\partial W_6} = Mz = 1(a_{H2}) = -0.2$$

iii.

$$f(z) = Mzf'(z) = M = 1 \tag{1}$$

$$\frac{\partial C}{\partial W_5} = -2(Y_i - a_o)(1)(a_{H1}) = -2.26 * 1 * -0.5 = 1.13$$

$$\frac{\partial C}{\partial W_6} = -2(Y_i - a_o)(1)(a_{H2}) = -2.26 * 1 * -0.2 = 0.452$$

iv. updated weights

$$W_5' = W_5 - \frac{\partial C}{\partial W_5} * \text{Learning Rate} = 0.3 - 1.13(0.1) = 0.187$$

$$W_6' = W_6 - \frac{\partial C}{\partial W_6} * \text{Learning Rate} = -0.1 - 0.452(0.1) = -0.1452$$