CSE 4088 – INTRODUCTION to MACHINE LEARNING

HW3 Report

Zeynep Kumralbaş 150115051

Gradient Descent

4) $\frac{\partial E}{\partial u} = 2(ue^v - 2ve^{-u})'(ue^v - 2ve^{-u}) = 2(e^v + 2ve^{-u})(ue^v - 2ve^{-u})$

Answer: e

5) Partial derivatives of E with respect to u and v are used to determine the gradient descent. Then, $\Delta w = -\mu * \nabla Ein(w_0)$ is calculated, weights are updated $w(t+1) = \Delta w + w(t)$. It took 10 iterations to fall the error below 10^{-14}.
Answer: d

6) u = 0.0447, v = 0.0240
Answer: e

7) The error is found 0.1398.
Answer: a

Logistic Regression

8) In one epoch, data points are selected randomly, and the w is updated based on the gradient of the chosen data point. After the w is updated for each data points, an epoch is finished. The w before the epoch and the w after an epoch are compared. When their difference becomes less than 0.01, the algorithm stops.
To find the E_{out}, 1000 new data points are created. Experiment is run 100 times, the average of the values is taken.
E_{out} is found 0.1026.
Answer: d

9) Number of epochs are found 342.33.

Answer: a

Regularization with Weight Decay

2) Non-linear transformation is performed, then linear regression is applied using

$X^{\dagger} = (X^T X)^{-1} X^T$ and $w = X^{\dagger} y$

E_{in} is found 0.0286, E_{out} is found 0.0840.

Answer: a

3) Linearization with weight decay formula is used.

$w = (X^T X + \lambda I)^{-1} X^T y$

E_{in} is found 0.0286, E_{out} is found 0.08.

Answer: d

4) k=3

E_{in} is found 0.3714, E_{out} is found 0.4360.

Answer: e

5) k=2, E_{out} is 0.2280.

k=1, E_{out} is 0.1280.

k=0, E_{out} is 0.0960.

k=-1, E_{out} is 0.060.

k=-2, E_{out} is 0.0840.

k=-1 has the smallest E_{out}.

Answer: d

6) Smallest E_{out} is 0.060.

Answer: b

Neural Networks

8)

- $w_{ij}^{l} x_i^{(l-1)}$is used in forward propagation.

  For $l = 1$;

$$i = 0 : w_{01}^1 x_0^0, \quad w_{02}^1 x_0^0, \quad w_{03}^1 x_0^0$$

$$i = 1 : w_{11}^1 x_1^0, \quad w_{12}^1 x_1^0, \quad w_{13}^1 x_1^0$$

$$i = 2 : w_{21}^1 x_2^0, \quad w_{22}^1 x_2^0, \quad w_{23}^1 x_2^0$$

$$i = 3 : w_{31}^1 x_3^0, \quad w_{32}^1 x_3^0, \quad w_{33}^1 x_3^0$$

$$i = 4 : w_{41}^1 x_4^0, \quad w_{42}^1 x_4^0, \quad w_{43}^1 x_4^0$$

$$i = 5 : w_{51}^1 x_5^0, \quad w_{52}^1 x_5^0, \quad w_{53}^1 x_5^0$$

We have 18 operations.

  For $l = 2$;

$$i = 0 : w_{01}^2 x_0^1$$

$$i = 1 : w_{11}^2 x_1^1$$

$$i = 2 : w_{21}^2 x_2^1$$

$$i = 3 : w_{31}^2 x_3^1$$

We have 4 operations.

  In total, we have 22 $w_{ij}^{l} x_i^{(l-1)}$operations.

- $w_{ij}^{l} \delta_i^{(l)}$is used in back propagation, $\delta_i^{(l-1)} = \left[1 - \left(x_i^{(l-1)}\right)^2\right] \sum_{j=1}^{d^{(l)}} w_{ij}^{l} \delta_i^{(l)}$.

  For $l = 2$;

$$\delta_i^{(1)} = \left[1 - \left(x_i^{(1)}\right)^2\right] \sum_{j=1}^{1} w_{ij}^2 \delta_i^{(2)}$$

$\delta_i^{(1)}$is found for $i = 1,2,3$. Therefore, we have 3 operations.

- $\frac{\partial e(w)}{w_{ij}^{(l)}} = x_i^{(l-1)} \delta_j^{(l)}$is used in back propagation.

  For $l = 2$;

$$x_0^1\delta_1^2, \quad x_1^1\delta_1^2, \quad x_2^1\delta_1^2, \quad x_3^1\delta_1^2$$

  For $l = 1$;

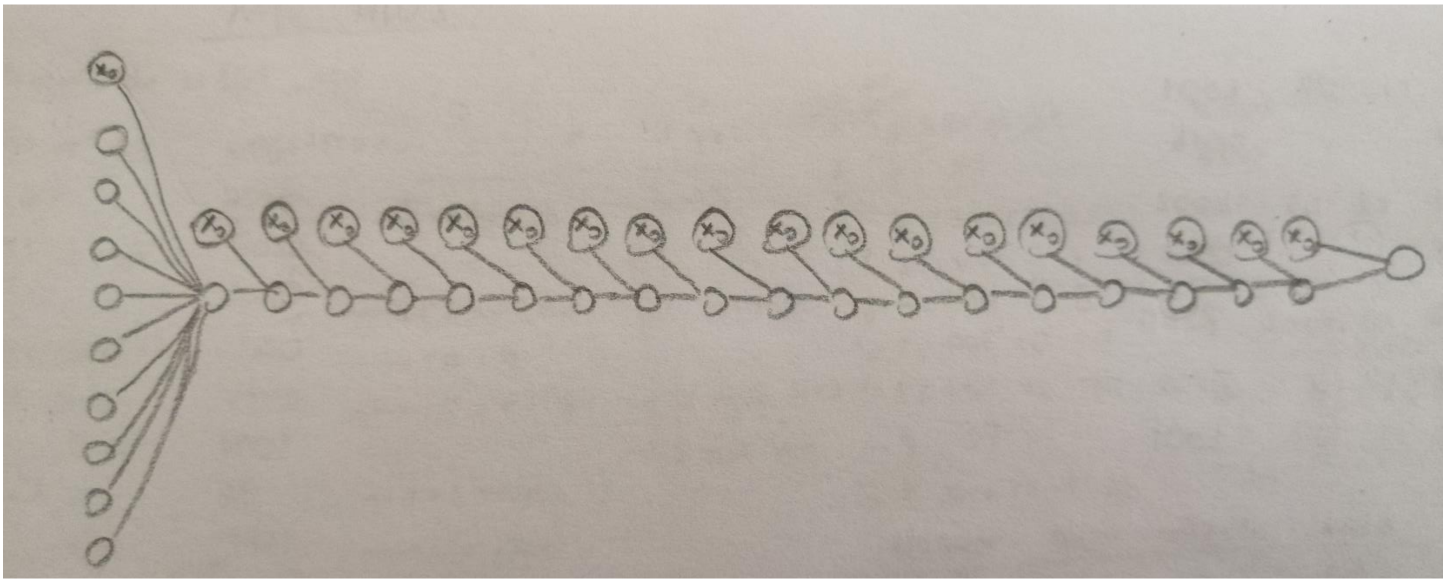
$$\begin{array}{cccccc} x_0^0\delta_1^1, & x_1^0\delta_1^1, & x_2^0\delta_1^1, & x_3^0\delta_1^1, & x_4^0\delta_1^1, & x_5^0\delta_1^1 \\ x_0^0\delta_2^1, & x_1^0\delta_2^1, & x_2^0\delta_2^1, & x_3^0\delta_2^1, & x_4^0\delta_2^1, & x_5^0\delta_2^1 \\ x_0^0\delta_3^1, & x_1^0\delta_3^1, & x_2^0\delta_3^1, & x_3^0\delta_3^1, & x_4^0\delta_3^1, & x_5^0\delta_3^1 \end{array}$$

  We have 22 operations.

In total, there are $22 + 3 + 22 = 47$ operations.

Answer: d

9) When the number of layers is increased, the total number of bias nodes(x_0) in layers are increased. Therefore, the number of weights is decreased. The minimum possible number of weights is achieved when we have neural network as in the figure:



Minimum number of weights: $y = 10 * 1 + (2 * 1) * 17 + 2 * 1 = 46$

Answer: a

10)

Assume $l = 1$

$$d^{(0)} = 10, d^{(1)} = 36, d^{(3)} = 1$$

The number of weights is $y = 10 * (36 - 1) + 36 * 1 = 386$

Assume $l = 2$

$$d^{(0)} = 10, d^{(1)} = a, d^{(2)} = b, d^{(3)} = 1$$

$$a + b = 36, \quad b = 36 - a$$

The number of weights is $y = 10 * (a - 1) + a * (b - 1) + b * 1$

$$y = 10 * (a - 1) + a * (35 - a) + (36 - a)$$

$$y = 10a - 10 + 35a - a^2 + 36 - a$$

$$y = -a^2 + 44a + 26a$$

The maximum number of weights is found: $y' = -2a + 44 = 0$

$$a = 22, b = 14$$

$$y = 10 * 21 + 22 * 13 + 14 * 1 = 510$$

Since there is no answer larger than 510, we do not need to try other possible number of levels. The maximum number of weights is 510.

Answer: e