Sohaib Syed
CS422
11/12/21
Dr. Gurbani

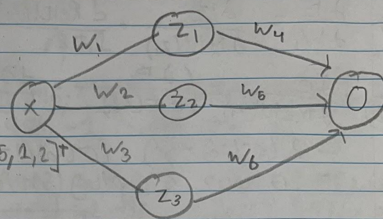
Homework 7

1.1

Homework 7

1.1 a)

$x = 4$
$y = 0$
$w_i : [1, 1, -1, 0.5, 1, 2]^T$



$\mathrm{Relu}(z_1) = 4(1) = 4 \quad z_1 0 = (4 \times 0.5) = 2$
$\mathrm{Relu}(z_2) = (4 \times 1) = 4 \quad z_2 0 = (4 \times 1) = 4$
$\mathrm{Relu}(z_3) = (4 \times -1) = 0 \quad z_3 0 = (0 \times 2) = 0$
$\sigma(2+4+0) = \boxed{.998} = z_0$

b) $[(.998 - 0)^2] = .996$

c) $\nabla E = \frac{\partial E}{\partial w_i} \forall = \left[\frac{\partial E}{\partial w_1} + \ldots + \frac{\partial E}{w_6}\right]^T$

$\frac{\partial E}{\partial w_4} = \frac{\partial E}{\partial \sigma(6)} \cdot \frac{d\sigma(6)}{d6} \cdot \frac{\partial 6}{\partial w_4}$
$= (2(\sigma(6) - 0)) \cdot [\sigma(6)(1 - \sigma(6))] \cdot \mathrm{relu}(4) = \boxed{.016 = w_4}$

$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial \sigma(6)} \cdot \frac{d\sigma(6)}{d(6)} \cdot \frac{\partial 6}{\partial w_5}$
$= 2(\sigma(6) - 0) \cdot [\sigma(6)(1 - \sigma(6))] \cdot \mathrm{relu}(4) = \boxed{.016 = w_5}$

$\frac{\partial E}{\partial w_6} = \frac{\partial E}{\partial \sigma(6)} \cdot \frac{d\sigma(6)}{d(6)} = \frac{\partial 6}{\partial w_6}$
$= 2(\sigma(6) - 0) \cdot [\sigma(6)(1 - \sigma(6))] \cdot \mathrm{relu}(-4) = \boxed{0 = w_6}$
$= \quad \cdot \quad \cdot \quad 0$

① $\frac{\partial E}{\partial W_1} = \frac{\partial E}{\partial ReLU(z_1)} \cdot \frac{\partial ReLU(z_1)}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1}$

② $\frac{\partial E}{\partial ReLU(z_1)} = \left(\frac{\partial E}{\partial \sigma(z_0)} \cdot \frac{\partial \sigma(z_0)}{\partial z_0} \cdot \frac{\partial z_0}{\partial relu(z_1)} \right)$

$= 2(\sigma(z_0) - y) \cdot [\sigma(z_0)(1-\sigma(z_0))] \cdot 0.5 =$

$= 0.002$ ← plug into $\underline{1}$

$.002(1) \cdot 4 = \boxed{.008 = W_1}$

$\frac{\partial E}{\partial W_2} = \frac{\partial E}{\partial ReLU(z_2)} \cdot \frac{\partial ReLU(z_2)}{\partial z_2} \cdot \frac{\partial z_2}{\partial W_2}$

$\frac{\partial E}{\partial ReLU(z_2)} = \left(\frac{\partial E}{\partial \sigma(z_0)} \cdot \frac{\partial \sigma(z_0)}{\partial(z_0)} \cdot \frac{\partial z_0}{\partial ReLU(z_2)} \right)$

$= 2(\sigma(z_0) - y) \cdot [\sigma(z_0)(1-\sigma(z_0))] \cdot 1 = .004$

$.004(1)(4) = \boxed{.016 = W_2}$

$\frac{\partial E}{\partial w_3} = \frac{\partial E}{\partial ReLU(z_3)} \cdot \frac{\partial ReLU(z_3)}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_3}$

$\frac{\partial E}{\partial ReLU(z_3)} = \left(\frac{\partial E}{\partial \sigma(z_0)} \cdot \frac{\partial \sigma(z_0)}{\partial(z_0)} \cdot \frac{\partial z_3}{\partial w_3} \right)$

$= 2(\sigma(z_0) - y) \cdot [\sigma(z_0)(1-\sigma(z_0))] \cdot 2 = 0.008$

$0.008(0)(4) = \boxed{0 = W_3}$

$$\boxed{\nabla E_w = [0.008, 0.016, 0, 0.016, 0.016, 0]^T}$$

d) $w = w - \eta \nabla E$

$\eta = 1$

$w = w - \nabla E$

$= [1, 1, -1, 0.5, 1, 2]^T - [0.008, 0.016, 0.000, 0.016, 0.016, 0]$

$= [.992, .984, -1, 0.484, .984, 2]$

$ReLu(z_1) = 4(.992) = 3.968$ $\quad z_1 0 = 3.968 \times .484 = 1.921$

$ReLu(z_2) = 4(.984) = 3.936$ $\quad z_2 0 = 3.936 \times .984 = 3.873$

$ReLu(z_3) = 4(-1) = 0$ $\quad z_3 0 = 0 \times 2 = 0$

$\sigma(1.921 + 3.873 + 0) = \boxed{.997}$

Loss: $(.997 - 0)^2 = \boxed{0.994}$

e) The loss value in d is less then the loss value in b. Even though it's a very small change, this was only one back prop pass. that appropriately adjusted weights so that the predicted value is closer to actual value.

1.2)

Chapter 7 Excerise 2.

2. Find all well-separated clusters in the set of points shown in **Figure 7.35**.

Exercise 6.

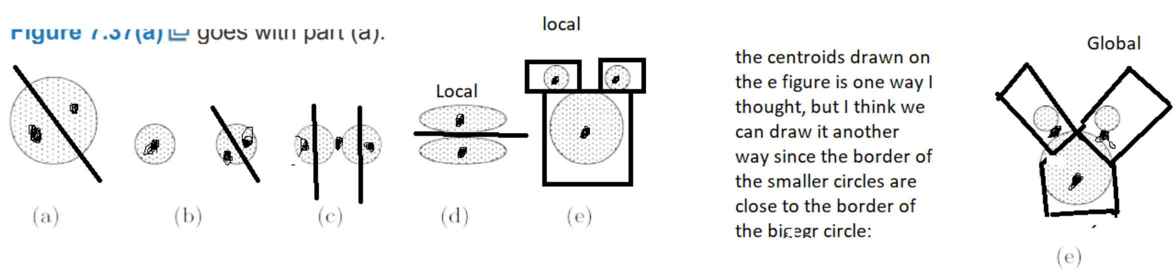
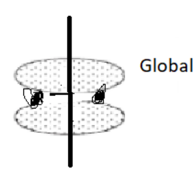


Figure 7.37.
Diagrams for **Exercise 6**.

D can also have a different way to get clustered:



a) There are infinite ways to partition the points if the points are uniformly distributed. All that has to be true is that the circle is bisected. The centroids will be a distance R/2 perpendicular to the bisected line. They will also be symmetrical to each other.

Exercise 11.

If the SSE is for one variable in all clusters, then that variable is likely to not be used to create further splits. It could also mean that that variable is in the correct cluster. If the SSE of the attribute is low for just one cluster, then that is helpful in being a characteristic of that one specific cluster. It means that that cluster has meaning when it comes to that attribute. If the SSE of the variable is high for all clusters then it could mean it is noise. If it is high just for one cluster then it really only means that the variable is not helpful in defining the cluster. Per variable SSE makes it easier to determine how to split clusters further to increase clusters in which the SSE's can be reduced

Exercise 12.

a) It sounds like an advantage of the leader algorithm is that the points and their distance to the leader point is computed only once. In the case that it is within the threshold set by the user, then it is added to the cluster of the leader point, or it is set as leader of a new cluster. Rather than every point's distance being repeatedly measured, the leader algorithm just creates a new cluster after the first miss. A disadvantage is that these clusters can not be directly predefined in the leader algorithm, such as in K-means.

b) A suggestion that I have is that rather than letting the user have complete control over the distance threshold, there could be a computation that takes place to first determine how spread out points are. This would however increase the number of computations which as of now is a strength of the leader algorithm.