22F CS-559 A HW 1

Jiangrui Zheng

December 1, 2022

1 Problem 1

1. Suppose initially we assign A_1, B_1, and C_2 as the center of each cluster.
   Cluster Assignment:

Distance(d^2)	A_1	B_1	C_2
A_1	0	13	5
A_2	25	18	20
A_3	72	25	41
B_1	13	0	2
C_1	65	52	58
C_2	5	2	0

   Then we can know minimums of distance between 3 centers. After the first iteration,
   Cluster 1: A_1
   Cluster 2: A_2, A_3, B_1, C_1
   Cluster 3: C_2

2. Cluster 1: $(2, 10)$
   Cluster 2: $(4, 4.75)$
   Cluster 3: $(4, 9)$

2 Problem 3

1. $X = (Gender = M, CarType = Family, ShirtSize = Large)$

   $P(X|Class = C0)$
   $= P(Gender = M|Class = C0) * P(CarType = Family|Class = C0) * P(ShirtSize = Large|Class = C0)$
   $= 6/10 * 1/10 * 2/10 = 0.012$

   $P(X|Class = C1)$
   $= P(Gender = M|Class = C1) * P(CarType = Family|Class = C1) * P(ShirtSize = Large|Class = C1)$
   $= 4/10 * 3/10 * 2/10 = 0.024$

   Since $P(C1) = P(C0) = 0.5, then P(X|Class = C1)P(C1) > P(X|Class = C1)P(C1)$ So $P(Class = C0|X) < P(Class = C1|X), Class = C1$.

2. Let X1 denote Car Type, X2 denote Gender, X3 denote Shirt Size.
   Bayesian Network for classification:

$$P\left(C \mid X_1, X_2, \ldots, X_n\right) = \frac{P\left(X_1, X_2, \ldots, X_n \mid C\right) P(C)}{P\left(X_1, X_2, \ldots, X_n\right)}$$

$$P\left(X_1 = Family, X_2 = M, X_3 = Large \mid C0\right) P(C0) = 6/10 * 1/6 * 0 * 5/10 = 0$$
$$P\left(X_1 = Family, X_2 = M, X_3 = Large \mid C1\right) P(C1) = 4/10 * 3/4 * 1/10 * 5/10 = 0.06$$

Since $0.06 > 0$, so $Class = C1$.

3 Problem 4

Computation graph:

$$X \longrightarrow \underbrace{W^1 X}_{z^1} \ \underbrace{\sigma(z^1)}_{a_1} \longrightarrow \underbrace{W^2 a_1}_{z^2} \ \underbrace{\sigma(z^2)}_{a_2} \longrightarrow \underbrace{W^3 a^2}_{z^3} \ \underbrace{\overset{S(z^3)}{\cancel{\sigma(z^3)}}}_{\hat{y}} \longrightarrow L(\hat{y}, y)$$

Back Progation:

$$\frac{\partial L}{\partial w^3} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z^3} \frac{\partial z^3}{\partial w^3}$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}\left[(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2\right] = -2(y_1 - \hat{y}_1) - 2(y_2 - \hat{y}_2)$$

$$\frac{\partial \hat{y}}{\partial z^3} = \frac{\partial}{\partial z^3}\left(\frac{e^{z_i^3}}{\sum_j e^{z_j^3}}\right) = \frac{\partial}{\partial z^3}\left(\frac{e^{z_1^3}}{\sum_j e^{z_j^3}}\right) + \frac{\partial}{\partial z^3}\left(\frac{e^{z_2^3}}{\sum_j e^{z_j^3}}\right) = (-\hat{y}_i \hat{y}_j) + (-\hat{y}_i \hat{y}_j) = -2 y_1 y_2$$

$$\frac{\partial z^3}{\partial w^3} = a^2$$

So $$\frac{\partial L}{\partial w^3} = -2a^2 y_1 y_2 \left[-2(y_1 - \hat{y}_1) - 2(y_2 - \hat{y}_2)\right] = 4a^2 y_1 y_2 \left[(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2)\right]$$

$$\frac{\partial L}{\partial w^2} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z^3} \frac{\partial z^3}{\partial a^2} \frac{\partial a^2}{\partial z^2} \frac{\partial z^2}{\partial w^2}$$

$$\frac{\partial z^3}{\partial a^2} = \frac{\partial}{\partial a^2}(w^3 a^2) = w^3$$

$$\frac{\partial a^2}{\partial z^2} = \frac{\partial(\sigma(z^2))}{\partial z^2} = \cancel{\frac{\partial}{\partial z^2}\left(\frac{1}{1+e^{-z^2}}\right)} = \sigma(z^2)(1 - \sigma(z^2)) = a^2(1-a^2)$$

$$\frac{\partial z^2}{\partial w^2} = a^1$$

So $$\frac{\partial L}{\partial w^2} = 4 y_1 y_2 \left[(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2)\right] w^3 a^1 a^2 (1-a^2)$$

$$\frac{\partial L}{\partial w^1} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z^3} \frac{\partial z^3}{\partial a^2} \frac{\partial a^2}{\partial z^2} \frac{\partial z^2}{\partial a^1} \frac{\partial a^1}{\partial z^1} \frac{\partial z^1}{\partial w^1}$$

$$\frac{\partial z^2}{\partial a_1} = \frac{\partial}{\partial a_1}(w^2 a^1) = w^2 \qquad \frac{\partial a^1}{\partial z^1} = \frac{\partial(\sigma(z^1))}{\sigma z^1} = a^1(1-a^1)$$

$$\frac{\partial z^1}{\partial w_1} = X$$

So $$\frac{\partial L}{\partial w^1} = 4 y_1 y_2 \left[(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2)\right] w^3 a^2 (1-a^2) w^2 a^1 (1-a^1) X$$

1.