AI HW-3

WRITTEN PART

SUBMITTED BY:
NAME: SANKUL RAWAT.
NET-ID: srawat5

A1.) Consider, each feature independent

Parameters → $P(Y)$ $P(x_i|Y)$ .....

where Y is label (class)
x_i feature vector

$$P(Y) \begin{cases} Y=0 \rightarrow 4/9 \\ Y=1 \rightarrow 5/9 \end{cases}$$

$$P(\text{Circle}|Y) \begin{cases} P(\text{circle}|Y=1) = 1/5 \\ P(\text{circle}|Y=0) = 1/4 \end{cases}$$

$$P(\text{Diamond}|Y) \begin{cases} P(\text{Diamond}|Y=1) = 1/5 \\ P(\text{Diamond}|Y=0) = 3/4 \end{cases}$$

$$P(\text{Square}|Y) \begin{cases} P(\text{Square}|Y=0) = 0/4 \\ P(\text{Square}|Y=1) = 3/5 \end{cases}$$

$$P(\text{Blue}|Y) \begin{cases} P(\text{Blue}|Y=0) = 1/4 \\ P(\text{Blue}|Y=1) = 2/5 \end{cases}$$

$$P(\text{Green}|Y) \begin{cases} P(\text{Green}|Y=0) = 3/4 \\ P(\text{Green}|Y=1) = 0/5 \end{cases}$$

$$P(\text{Red}|Y) \begin{cases} P(R|Y=0) = 0/4 \\ P(R|Y=1) = 3/5 \end{cases}$$

A2.) For a test sample (shape, color) to give undefined posterior label probability it should give :-

$$P(\text{color}, \text{shape}, Y) = 0$$

The only combination which does so is (square, green)

$$P(\text{square}, \text{green}, Y=0) = \frac{0}{4} \times \frac{3}{4} = 0$$

$$P(\text{square}, \text{green}, Y=1) = \frac{0}{5} \times \frac{3}{5} = 0$$

A3.) Parameters :-

$$P(Y) = \begin{cases} P(Y=0) = 5/11 \\ P(Y=1) = 6/11 \end{cases}$$

$$P(\text{circle} \mid Y) = \begin{cases} P(C \mid Y=0) = 2/7 \\ P(C \mid Y=1) = 2/8 \end{cases}$$

$$P(\text{diamond} \mid Y) = \begin{cases} P(D \mid Y=0) = 4/7 \\ P(D \mid Y=1) = 2/8 \end{cases}$$

$$P(\text{Square} \mid Y) = \begin{cases} P(S \mid Y=0) = 1/7 \\ P(S \mid Y=1) = 4/8 \end{cases}$$

$$P(\text{Blue} \mid Y) = \begin{cases} P(B \mid Y=0) = 2/7 \\ P(B \mid Y=1) = 3/8 \end{cases}$$

$$P(\text{Red} \mid Y) = \begin{cases} P(R \mid Y=0) = 1/7 \\ P(R \mid Y=1) = 4/8 \end{cases}$$

$$P(\text{Green} \mid Y) = \begin{cases} P(G \mid Y=0) = 4/7 \\ P(G \mid Y=1) = 1/8 \end{cases}$$

A4.) $P(Y=1 \mid \text{Green}, \text{Square}) = P(\text{Green}, \text{Square} \mid Y=1) \times P(Y=1)$

$$\Rightarrow \frac{\frac{1}{8} \times \frac{4}{8} \times \frac{6}{11}}{\frac{1}{8} \times \frac{4}{8} \times \frac{6}{11} + \frac{5}{11} \times \frac{4}{7} \times \frac{1}{7}}$$

$$\Rightarrow \frac{0.034}{0.034 + 0.037} \Rightarrow 0.47$$

$$P(Y=1 \mid \text{Green}, \text{Square}) = 1 - P(\text{Green}, \text{Square} \mid Y=1)$$

$$= 1 - 0.47$$

$$= 0.53$$

A5.) $\alpha_0, \alpha_1, \alpha_2$

↳ bias term

Initial weight = (1, 0, 0)
$\alpha_0\ \alpha_1\ \alpha_2$

Assume $\alpha_0 = 1, x_0 = 1$

Given :- if $\alpha_0 x_0 + \alpha_1 x_1 + \alpha_2 x_2 > 0$ then +

else −

Iteration - 1

(1, 4, +) → $1\times1 + 0\times1 + 0\times4 = 1 \Rightarrow +$

(3, 2, +) → $1\times1 + 0\times3 + 0\times2 = 1 \Rightarrow +$

(1, 2, −) → $1\times1 + 0\times1 + 0\times2 = 1 \Rightarrow +$ but should be −

New weight = old weight + actual label × Feature vector (x_0, x_1, x_2)

= (1, 0, 0) + (−1) × (1, 1, 2)

= (0, −1, −2)

(2, 1, −) → $0\times1 + (-1)\times2 + (-2)\times1 = -4$

Iteration - 2

(1, 4, +) → $0\times1 + (-1)\times1 + (-2)\times4 = -9 \rightarrow -$ should be +

New weight = [1, 0, 2)

(3, 2, +) → $1\times1 + 0\times3 + 2\times2 = 5 \rightarrow +$

(1, 2, −) → $1\times1 + 0\times1 + 2\times2 = 5 \rightarrow +$ should be −

New weight = (0, −1, 0).

(2, 1, −) → $0\times1 + (-1)\times2 + 0\times1 = -2 \rightarrow -$

Iteration - 3 [Running in sequence but will only mention misclassified points & new weights]

Points misclassified

(1, 4, +) → −1 → − should be +

New weight = (1, 0, 4)

(1, 2, −) → 9 → + should be −

New weight = (0, −1, 2)

Iteration – 4

Points misclassified

(1, 2, -) → 3 → + should be -

New weight = (-1, -2, 0)

Iteration – 5

Points misclassified

(1, 4, +) → -3 → - should be +

New weight = (0, -1, 4)

(1, 2, -) → 7 → + should be -

New weight = (-1, -2, 2)

Iteration – 6

(3, 2, +) → -3 → - should be +

New weight = (0, 1, 4)

(1, 2, -) → 9 → + should be -

New weight = (-1, 0, 2)

(2, 1, -) → 1 → + should be -

New weight = (-2, -2, 1)

Iteration – 7

(1, 4, +) → 0 → - should be +

New weight = (-1, -1, 5)

(1, 2, -) → 8 → + should be -

New weight = (-2, -2, 3)

Iteration - 8

(3, 2, +) → −5 → − should be +

New weight = (−1, 1, 5)

(1, 2, −) → 10 → + should be −

New weight = (−2, 0, 3)

(2, 1, −) → 1 → + should be −

New weight = (−3, −2, 2)

Iteration - 9

(3, 2, +) → −5 → − should be +

New weight = (−2, 1, 4)

(1, 2, −) → 7 → + should be −

New weight = (−3, 0, 2)

Iteration - 10

(1, 2, −) → 1 → + should be −

New weight = (−4, −1, 0)

Iteration - 11

(1, 4, +) → −5 → − should be +

New weight = (−3, 0, 4)

(1, 2, −) → 5 → + should be −

New weight = (−4, −1, 2)

Iteration - 12

(3, 2, +) → −3 → − should be +

New weight = (−3, 2, 4)

(1, 2, −) → 7 → + should be −

New weight = (−4, 1, 2)

Iteration - 13

(1, 2, −) → 1 → + should be −

New Wt. = (−5, 0, 0)

Iteration − 14

(1, 4, +) → −5 → − should be +

New Wt. = (−4, 1, 4)

(1, 2, −) → 5 → + should be −

New Wt. = (−5, 0, 2)

Iteration − 15

(3, 2, −) → −1 → − should be +

New Wt. = (−4, 3, 4)

(1, 2, −) → 1 → + should be −

New Wt. = (−5, 2, 2)

(2, 1, −) → 1 → + should be −

New Wt. = (−6, 0, 1)

Iteration − 16

(1, 4, +) → −2 → − should be +

New Wt. = (−5, 1, 5)

(1, 2, −) → 6 → + should be −

New Wt. = (−6, 0, 3)

Iteration − 17

(3, 2, +) → 0 → − should be +

New Wt. = (−5, 3, 5)

(1, 2, −) → 8 → + should be −

New Wt. = (−6, 2, 3)

(2, 1, −) → 1 → + should be −

New Wt. = (−7, 0, 2)

Iteration -18

(3, 2, +) → -3 → − should +

New wt. = (-6, 3, 4)

(1, 2, -) → 5 → + should be -

New wt. = (-7, 2, 2)

Iteration -19

No misclassification.

Final converged weights = (-7, 2, 2)

A6.) let data points be :-

1010, 1020, 1030, 1100, 1110, 1120, 2000, 2010, 2020, 2500, 2510, 2520

If we take 1010, 1020, 1100, 2000 as initial seeds then we get a non-optimal clusters.

Iteration -1

Cluster 1 → 1010 , mean = 1010

Cluster 2 → 1020, 1030 , mean = 1025

Cluster 3 → 1100, 1110, 1120 , mean = 1110

Cluster 4 → 2000, 2010, 2020, 2500, 2510, 2520, mean = 2260

New centroids = 1010, 1025, 1110, 2260

In iteration -2 the centroids remain same Hence, the algorithm converges.

The reason that it converges to non-optimal is as K-means is sensitive to initial seeds.

i.e if you choose them wrong the clusters might not end up being optimal.

An example of optimal cluster :-

Let initial seeds be:

1020, 1110, 2010, 2510

Cluster 1 → 1010, 1020, 1030, mean = 1020

Cluster 2 → 1100, 1110, 1120, mean = 1110

Cluster 3 → 2000, 2010, 2020, mean = 2010

Cluster 4 → 2500, 2510, 2520, mean = 2510

In iteration -2 the centroids would remain the same. Hence, algo converges on 1020, 1110, 2010, 2510 and the clusters end up being optimal.