CS584 – Machine Learning
Homework #2 Writing
Submitted by: Abdurakhmon Urazboev (aurazboev@hawk.iit.edu)

1. If we are rolling 2 dice, each of the dice have 6 sides numbered from 1 to 6. So, we can have $6 \times 6 = 36$ possible outcomes. We need to know what permutation will result in sum greater than 8.
   Permutations that give sum greater than 8 is as follows:
   6+3, 6+4, 6+5, 6+6, 5+6, 5+5, 5+4, 4+6, 4+5, 3+6
   So, in total we have 10 outcomes that will give us result greater than 10.
   a) If we do not have any information, the probability of us getting greater than 8 is equal to the number of desirable outcomes over total number of outcomes: $\frac{10}{36} = \frac{5}{18}$.
   b) If we know one dice displays number 5, in order to find out the probability of getting number greater than 8 we use the same approach. Now since 1 dice is already known, we can calculate the probability for only one dice. We can have 6 different outcomes and 5+6, 5+5, 5+4 these are the outcomes that we need. So, the probability of getting a number greater than 8 is $\frac{3}{6} = \frac{1}{2}$
   c) Now we know that 1 dice shows number greater than 2. This will reduce the total number of possible outcomes. Now one dice can have only 4 outcomes while the other has 6 possible outcomes. So, total number of outcomes is $4 \times 6 = 20$. Number of desired outcomes remains the same as none of them has dice that display number less than 3. So, the probability of getting a sum greater than 8 in this case is $\frac{10}{20} = \frac{1}{2}$

2. To answer this question, we will use Bayes' rule that we learned in the lecture.
   Which says $P(cause|effect) = \frac{P(effect|cause)*P(cause)}{P(effect)}$ Now, let's define all probabilities that we have. We have $P(effect|cause)$ given in terms of the test being positive if certain type of disease is there.
   $P(+|C1) = 0.85, P(+|C2) = 0.4,\ P(+, C3) = 0.3.$
   Equal probability is assumed for each type of disease so $P(cause)$ is same for all of them. $P(C1) = P(C2) = P(C3) = \frac{1}{3}$. We are not however given explicitly of probability of getting positive results so we should calculate them by summing the probability for each type of disease. We get $P(+) = \sum_1^3 P(+|C_n) \times P(C_n) = 0.85 \times \frac{1}{3} + 0.4 \times \frac{1}{3} + 0.3 \times \frac{1}{3} = \frac{1.55}{3} = \frac{31}{60}$.
   Now we have everything that we need to calculate probabilities for patient to have particular type of disease.
   $$P(C1|+) = \frac{0.85 \times \frac{1}{3}}{\frac{31}{60}} = \frac{20 \times 0.85}{30} = \frac{17}{30}$$

$$P(C1|+) = \frac{0.4 \times \frac{1}{3}}{\frac{31}{60}} = \frac{20 \times 0.4}{30} = \frac{4}{15}$$

$$P(C1|+) = \frac{0.3 \times \frac{1}{3}}{\frac{31}{60}} = \frac{20 \times 0.3}{30} = \frac{1}{5} = 0.2$$

3. We are to conduct gradient descent algorithm for linear regression model with intercept which is $y' = w_0 + w_1 x$. As we are given initial values as 0.1 and 0.1 along with learning rate of 0.5 and the model will be $y' = 0.1 + 0.1x$. Let's start with calculating initial MSE:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - {y'}_i)^2$$

a) ${y'}_1 = 0.1 + 0.1 \times 3 = 0.4$
b) ${y'}_2 = 0.1 + 0.1 \times 4 = 0.5$
c) ${y'}_3 = 0.1 + 0.1 \times 5 = 0.6$
d) ${y'}_4 = 0.1 + 0.1 \times 6 = 0.7$

$$MSE = \frac{1}{4}[(5 - 0.4)^2 + (6 - 0.5)^2 + (8 - 0.6)^2 + (8 - 0.7)^2]$$
$$= \frac{1}{4}(21.16 + 30.25 + 54.76 + 53.29) = \frac{159.46}{4} = 39.865$$

Gradient Descent Update calculations:

$\delta w_0 = \frac{-2}{N}\sum_{i=1}^{N}(y_i - {y'}_i)$, $\delta w_1 = \frac{-2}{N}\sum_{i=1}^{N}(y_i - {y'}_i) \times x$

First iteration:

$$\delta w_0 = \frac{-2}{4}(4.6 + 5.5 + 7.4 + 7.3) = \frac{-2 \times 24.8}{4} = -12.4$$
$$\delta w_1 = \frac{-2}{4}(4.6 * 3 + 5.5 * 4 + 7.4 * 5 + 7.3 * 6) = \frac{-2}{4}(13.8 + 22 + 37 + 43.8)$$
$$= \frac{-116.6}{2} = -58.3$$

$w_0' = w_0 - \alpha \times \delta w_0 = 0.1 - 0.5 \times (-12.4) = 0.1 + 6.2 = 6.3$
$w_1' = w_1 - \alpha \times \delta w_1 = 0.1 - 0.5 \times (-58.3) = 0.1 + 29.15 = 29.25$

So, our updated weights after first iteration are $w_0' = 6.3, w_1' = 29.25$ and the updated model is $y' = 6.3 + 29.25x$

Let's now calculate MSE for this model:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - {y'}_i)^2$$

e) ${y'}_1 = 6.3 + 29.25 \times 3 = 94.05$
f) ${y'}_2 = 6.3 + 29.25 \times 4 = 123.3$
g) ${y'}_3 = 6.3 + 29.25 \times 5 = 152.55$
h) ${y'}_4 = 6.3 + 29.25 \times 6 = 181.8$

$$MSE = \frac{1}{4}[(5 - 94.05\ \ )^2 + (6 - 123.3)^2 + (8 - 152.55)^2 + (8 - 181.8)^2]$$
$$= \frac{1}{4}(7929.9025 + 13759.29 + 20894.7025 + 30206.44)$$
$$= \frac{72790.335}{4} = 18197.58375$$

Update calculations for second iteration:

$$\delta w'_0 = \frac{-2}{4}(-89.05 - 117.3 - 144.55 - 173.8) = \frac{-2 \times (-524.7)}{4} = 262.35$$
$$\delta w'_1 = \frac{-2}{4}(-89.05 * 3 - 117.3 * 4 - 144.55 * 5 - 173.8 * 6)$$
$$= \frac{-2}{4}(-267.15 - 469.2 - 722.75 - 1042.8) = \frac{2501.9}{2} = 1250.95$$

$$w_0'' = w'_0 - \alpha \times \delta w'_0 = 6.3 - 0.5 \times 262.35 = -124.875$$
$$w_1'' = w'_1 - \alpha \times \delta w'_1 = 29.25 - 0.5 \times 1250.95 = -596.225$$

So, the model after the second iteration will be $y'' = -124.875 - 596.225x$

Let's now calculate MSE for this model:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - y'_i)^2$$

i) $y'_1 = -124.875 - 596.225 \times 3 = -1913.55$
j) $y'_2 = -124.875 - 596.225 \times 4 = -2509.775$
k) $y'_3 = -124.875 - 596.225 \times 5 = -3106$
l) $y'_4 = -124.875 - 596.225 \times 6 = -3702.225$

$$MSE = \frac{1}{4}[(5 + 1913.55\ )^2 + (6 + 2509.775)^2 + (8 + 3106)^2 + (8 + 3702.225)^2]$$
$$= \frac{1}{4}(3680834.1025 + 6329123.85063 + 9696996$$
$$+ 13765769.5506) = \frac{33472723.5037}{4} = 8368180.87593$$

4. In multi-label classification we use the following formula:
$f_{CE} = -\sum_{i=1}^{n}\sum_{k=1}^{c} y_k^{(i)} log y'^{(i)}_k$ shown on the lecture.
We are given model outputs:

Green	Blue	Red
0.6	0.3	0.1
0.4	0.5	0.1
0.5	0.3	0.2
0.3	0.3	0.4
0.3	0.5	0.2

0.1	0.8	0.1
0.2	0.7	0.1
0.3	0.2	0.5
0.4	0.3	0.3
0.2	0.5	0.3

And we are given Training dataset with labels which we convert into one hot format with labels:

Green	Blue	Red
1	0	0
0	1	0
0	1	0
0	0	1
1	0	0
1	0	0
0	1	0
1	0	0
1	0	0
1	0	0

Now we calculate the cross entropy and take average $-\frac{1}{10}\left(log2(0.6) + log2(0.5) + log2(0.3) + log2(0.4) + log2(0.3) + log2(0.1) + log2(0.7) + log2(0.3) + log2(0.4) + log2(0.2)\right) = -\frac{1}{10} \times (-15.75) = 1.575.$

The confusion matrix for this problem will be what $actual\ x\ predicted$ so we will have 3x3 matrix.

	Actual			
predicted		Green	Blue	Red
	Green	2	2	0
	Blue	3	1	0
	Red	1	0	1