Q1 Problem Classification
10 Points

Explain whether each scenario is a classification or regression problem, and provide N and d (d is the number of features).

Q1.1 Firms
5 Points

We collect a set of data on the top 500 firms in the US. For each firm we record profit, number of employees, industry, and the CEO salary. We are interested in understanding which factors affect CEO salary.

Classification

Regression

Explanation:

The value predicted, CEO salaries, will be within a range of continous real value output values which is why it would be a regression problem.

N:

500

d:

3

Q1.2 Products
5 Points

We are considering launching a new product and wish to know whether it will be a *success* or *failure*. We collect data on 20 similar products that were previously launched. For each product we have recorded whether it was a success or failure, price charged for the product, marketing budget, competition price, and ten other variables.

Classification

Regression

Explanation:

The value predicted, success or failure are binary values which is why it would be a classification problem.

N:

20

d:

13

Q2 ML Applications
10 Points

Think of real-life applications for machine learning

Q2.1 Classification
5 Points

Describe three real-life applications in which *classification* might be useful. Describe the target, as well as the features. Explain the application (inference/prediction).

Example 1:

A real-life application in which classification might be useful is in the field of healthcare, specifically in diagnosing cancer.

Target: Diagnosing cancer

Features:
patient symptoms, medical history, and results from medical tests (such as biopsy results, imaging studies, and blood tests).

Example 2:

A real-life application in which classification might be useful is in the field of robot vision, specifically in self driving cars.

Target: Classify images into stop signs.

Features:
This can include information such as the objects present in the image, their shapes, colors, and positions, information about the location, lighting conditions, and camera angle of the image.

Example 3:

A real-life application in which classification might be useful is in the field of email spam filtering based on the contents of the email.

Target: Emails in a user's inbox and classify them as spam or not spam.

Features: information such as the sender, the subject line, the content of the email, and any attached files, known spam emails

Q2.2 Regression
5 Points

Describe three real-life applications in which *regression* might be useful. Describe the target, as well as the features. Explain the application (inference/prediction).

Example 1:

A real-life application in which regression might be useful is in the field of stock market price prediction.

Target: Predict the price of a stock at a future time.

Features: information such as the financials of the company, the company's competitors' financials, company's industry trends, the team

Example 2:

A real-life application in which regression might be useful is in the field of predicting energy consumption of a neighborhood.

Target: Predict the amount of energy used by a neighborhood over a period of time.

Features: information such as the number of people in that neighborhood, past consumption data, the weather, the time of day, the day of the week, the season.

Example 3:

A real-life application in which regression might be useful is in the field of predicting the life of a battery.

Target: Predict the life of a battery.

Features: information such as the Type of battery (e.g. Lithium-ion, Lead-Acid, etc.), Capacity (mAh), Voltage (V), Current draw (A), temperature, Age of the battery

Q3
15 Points

A university admissions office wants to predict the success of students based on their application material. They have access to past student records as training data.

Q3.1 Target Variable
3 Points

To formulate this as a supervised learning problem, identify a possible target variable. This should be some variable that measures success in a meaningful way and can be easily collected (in an automated manner) by the university. There is no one correct answer to this problem.

Salary of the student after graduation.

Q3.2 Continuous or Discrete
4 Points

Is the target variable continuous or discrete-valued?

Continuous

Discrete

Q3.3 Predictor
4 Points

State at least one possible variable that can act as the predictor for the target variable you chose in Q3.1.

income of the student's family

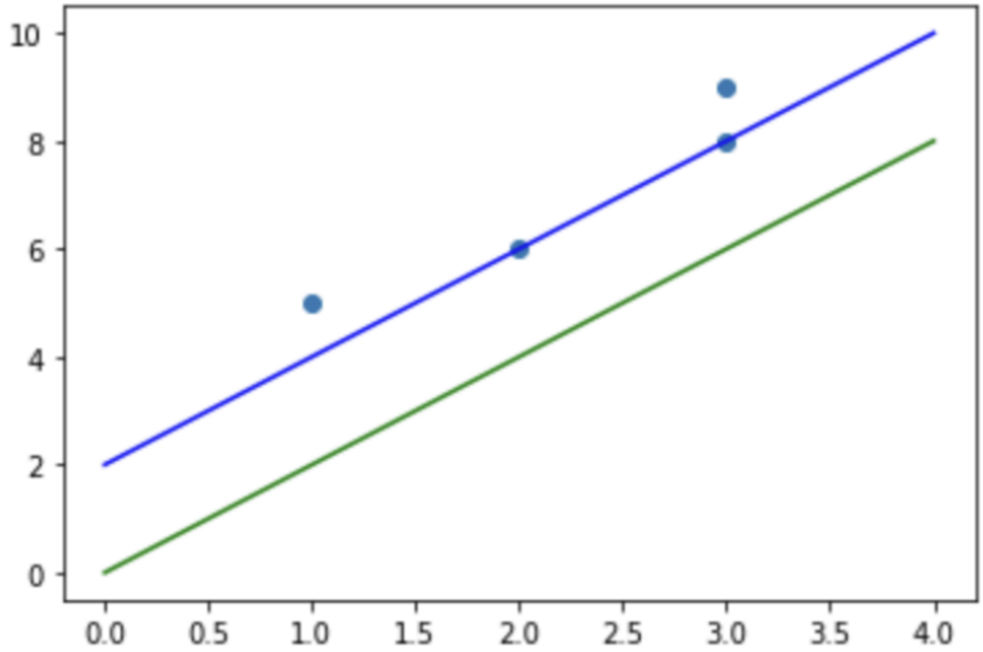
Q3.4 Linear?
4 Points

Before looking at the data, would a linear model for the data be reasonable? If so, what sign do you expect the slope to be?

I would not expect there to be a linear relationship. A student from a high income family may have many relationships in their chosen industry, been brought up with good education and grades, which may all contribute to a linear relationship, but there will be several cases where a student from a low income family gets very high salary because of their struggles/innate strengths inspite of not getting the advantages. There would be a positive slope in the graph.

Q4 Continued
30 Points

Consider the data $(1,5),(2,6),(3,8),(3,9)$ and regression lines: $y = 2x_1$ (the green line), and $y = 2x_1 + 2$ (the blue line). (Note here $x = [x_1]$)



Q4.1 Squared Error
6 Points

What is the squared error of each of the points with respect to the line $y = 2x_1$?

Squarred error of 2x_1 = (2 - 5)^2 + (4 - 6)^2 + (6 - 8)^2 + (6 - 9)^2 = 26

Q4.2 Gradient Contribution
6 Points

The gradient of our cost function includes a sum over contributions of individual points. We could calculate the individual contributions separately.

The gradient for a single $(x^{(i)}, y^{(i)})$ point is:

$$\begin{bmatrix} (w_0 + w_1 * x_1^{(i)} - y^{(i)}) \\ (w_0 + w_1 * x_1^{(i)} - y^{(i)})x_1^{(i)} \end{bmatrix}$$

For the line $y = 2x_1$ (the green line), what is the gradient contribution for each of the four examples?

[

(0 + 2 * 1 - 5)
(0 + 2 * 1 - 5) * 1
] = [
-3
-3
],

[
(0 + 2 * 2 - 6)
(0 + 2 * 2 - 6) * 2
] = [
-2
-4
],

[
(0 + 2 * 3 - 8)
(0 + 2 * 3 - 8) * 3
] = [
-2
-6
],

[
(0 + 2 * 3 - 9)
(0 + 2 * 3 - 9) * 3
] = [
-3
-9
]

Q4.3 Squared Error 2
6 Points

What is the squared error of each of the points with respect to the line $y = 2x_1 + 2$ (the blue line)?

Squarred error of 2x_1 + 2 = (4 - 5)^2 + (6 - 6)^2 + (8- 8)^2 + (8 - 9)^2 = 2

Q4.4 Gradient Contribution 2
6 Points

What is the gradient contribution for each of the four examples to the line $y = 2x_1 + 2$ (the blue line)?

```
[
    (2 + 2 * 1 - 5)
    (2 + 2 * 1 - 5) * 1
] = [
    -1
    -1
],

[
    (2 + 2 * 2 - 6)
    (2 + 2 * 2 - 6) * 2
] = [
    0
    0
],

[
    (2 + 2 * 3 - 8)
    (2 + 2 * 3 - 8) * 3
] = [
    0
    0
],

[
    (2 + 2 * 3 - 9)
    (2 + 2 * 3 - 9) * 3
] = [
    -1
    -3
]
```

Q4.5 RSS
2 Points

Which line has a smaller RSS?

$y = 2x_1$ (green line)

$y = 2x_1 + 2$ (blue line)

Q4.6 RSS 2
4 Points

Would it be possible for a different line to have a smaller RSS?

Yes

No

Why?

We use a linear relationship here to find the best fit line between the points, however, it is possible to use a polynomial relationship curve which would lead to a smaller RSS and could better fit the points.

Q5
26 Points

Given the following data: $((x_1, x_2)^T, y)$:
$((0,0)^T, 1), ((0,1)^T, 4), ((1,0)^T, 3), ((1,1)^T, 7))$

For this problem, whenever you need to calculate a value:
1) Provide the formulas using the numbers given in the problem
2) You can actually calculate the values using program or a calculator. Just make sure you could do this by hand if you need to in an exam. These instructions are true for any future homework assignment unless otherwise specified.

Q5.1 Design Matrix
1 Point

Create the design matrix X (include the column of 1's)

[
  [1, 0,   0],

[1, 0,   1],
[1, 1,   0],
[1, 1,   1]
]

Q5.2 Target Vector
1 Point

Create the target vector y

[
[1],
[4],
[3],
[7],
]

Q5.3 Closed Form Solution
2 Points

Write out the closed form solution for computing w that minimizes $RSS(w)$

To find the globally optimal solution (closed form solution), we use the expression
w_in = (X_T * X)^(-1) * X_T * y

Q5.4 w's
4 Points

Determine the w_0, w_1, w_2 that minimizes $RSS(w)$

w0 = 0.75
w1 = 2.5
w2 = 3.5

Q5.5 RSS
4 Points

Compute RSS

RSS = 0.25
sum((y_i - (w0 + w1 * x_i1 + w2 * x_i2))^2) = 0.25

Q5.6 TSS
4 Points

Compute TSS

TSS = 18.5
sum(((w0 + w1 * x_i1 + w2 * x_i2) - y_avg)^2)

mean = y_avg = (1+4+3+7)/4 = 3.75

Q5.7 R^2
4 Points

Compute R^2

R^2 = 1 - TSS/RSS = 1 - 0.25/18.75 = 0.987

Q5.8 Portion of Variance?
2 Points

What portion of variance in y is explained by x?

Based on our R^2 analysis, it is found that 98.7% of the variability in the dependent
variable y is accounted for or explained by the independent variable x. This is considered
a very strong relationship between x and y and indicates that x is a very good predictor of y.

Q5.9 Prediction
4 Points

Predict the value of $x^T = (0.5, 0.5)$ using the values of w computed in question 5.4

y_pred = 0.75 + 2.5 * 0.5 + 3.5 * 0.5 = 3.75

Q6 Different Alphas
10 Points

For the following function: $f_1(w_0, w_1) = (w_0 + 2w_1 - 4)^2 + (w_0 + 3w_1 - 3)^2$

Q6.1 Gradient
4 Points

Determine the gradient $\nabla f(w_0, w_1)$

[
 2 * (w_0 + 2*w_1 - 4) + 2 * (w_0 + 3*w_1 - 3)
 2 * 2 * (w_0 + 2*w_1 - 4) + 2 * 3 * (w_0 + 3*w_1 - 3)
]

Q6.2 Gradient Descent
6 Points

Run the gradient descent algorithm for num_iters = 10 iterations (you can use your computer to perform the calculations) where you try different learning rates. For each start with $(w_0, w_1) = (0, 0)$.

Report the value of w_0, w_1, and $f(w_0, w_1)$ at the end of each step. On one graph, plot the points (w_0, w_1) at every iteration.

Evaluate (briefly, in one sentence) how each learning rate contributed or did not contribute to finding a new assignment to the parameters that decreased the value of the function.

$\alpha = 0.06$

We see that both weights are substantially being changes in values after each iteration (as shown by the gaps between the blue points). This means that we are learning huge amounts from each iteration which would be a good idea since our iterations are very few (only 10). But we also see an overshoot in w0 which was then recovered on the next iteration.

=================================

alpha: 0.06
iteration_num 0 w0: 0.84 w1: 2.04 f_w: 16.528
iteration_num 1 w0: 0.25440000000000007 w1: 0.3936000000000004 f_w: 11.200729599999992
iteration_num 2 w0: 0.7971839999999998 w1: 1.6669439999999995 f_w: 7.846073405439993
iteration_num 3 w0: 0.4456934400000001 w1: 0.6282009600000005 f_w: 5.728868753539064
iteration_num 4 w0: 0.8018064383999997 w1: 1.4207913983999996 f_w: 4.388012925700533
iteration_num 5 w0: 0.5968980541439999 w1: 0.7632729538560004 f_w: 3.5342956214805055
iteration_num 6 w0: 0.8356787488358397 w1: 1.2544283133542398 f_w: 2.986318752328914
iteration_num 7 w0: 0.7224588611026943 w1: 0.8361128952201219 f_w: 2.630304875583665
iteration_num 8 w0: 0.8874009973059745 w1: 1.1383014620151153 f_w: 2.3948887550470705
iteration_num 9 w0: 0.8314438807434715 w1: 0.8701105828879508 f_w: 2.2353064460727423

$\alpha = 0.001$

Clearly, the number of iterations were insufficient for this alpha case because the weights did not reach their optimal values in 10 reductions and learned the least out of all (as evidenced by their low values in w0 and w1).

=================================

alpha: 0.001

iteration_num 0 w0: 0.014 w1: 0.034 f_w: 23.66818
iteration_num 1 w0: 0.027604 w1: 0.06697600000000001 f_w: 22.41468721616
iteration_num 2 w0: 0.040823823999999995 w1: 0.09895858400000002 f_w: 21.234912872891726
iteration_num 3 w0: 0.053670942863999996 w1: 0.129977422576 f_w: 20.124519375676186
iteration_num 4 w0: 0.066156484866784 w1: 0.160061300160384 f_w: 19.07942435472743
iteration_num 5 w0: 0.07829124592571302 w1: 0.18923814150754617 f_w: 18.09578564752956
iteration_num 6 w0: 0.0900856995269347 w1: 0.21753503736909283 f_w: 17.169987165003644
iteration_num 7 w0: 0.10155000635513603 w1: 0.24497826940222706 f_w: 16.298625589311527
iteration_num 8 w0: 0.11269402363569321 w1: 0.2715933343342178 f_w: 15.478497854362981
iteration_num 9 w0: 0.12352731419780826 w1: 0.2974049674051712 f_w: 14.706589362971709

$\alpha = 0.03$

This learning rate seems optimal with the number of iterations as opposed to the other options because it initially takes big steps in learning and towards the end, it takes small steps to indicate that the optimal values have been reached for w0 and w1

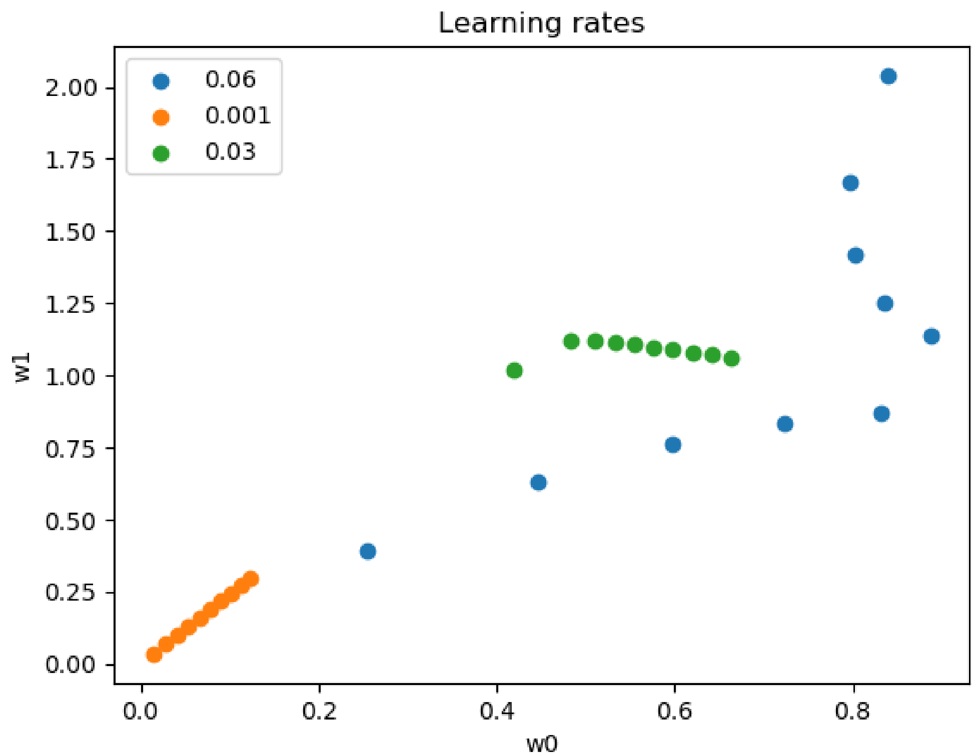
==================================

alpha: 0.03
iteration_num 0 w0: 0.42 w1: 1.02 f_w: 2.602
iteration_num 1 w0: 0.4836 w1: 1.1184 f_w: 2.3409615999999995
iteration_num 2 w0: 0.510048 w1: 1.120968 f_w: 2.3195891305600003
iteration_num 3 w0: 0.53255184 w1: 1.11359856 f_w: 2.3009585329112325
iteration_num 4 w0: 0.5545660511999999 w1: 1.1052261311999998 f_w: 2.282505159448647
iteration_num 5 w0: 0.576450285696 w1: 1.096779933504 f_w: 2.2642000779997096

iteration_num 6 w0: 0.5982422713612799 w1: 1.08835649966208 f_w: 2.2460418016419528
iteration_num 7 w0: 0.6199462488993023 w1: 1.0799657485172736 f_w: 2.2280291498173828
iteration_num 8 w0: 0.6415629744762039 w1: 1.0716085900040095 f_w: 2.21016095462152
iteration_num 9 w0: 0.6630928405378566 w1: 1.063284997458021 f_w: 2.1924360575506365

Graph results of all three learning rates together, and ensure they can be differentiated from each other. Upload graph:

▾ q6.2_graph.png Download



Q7 Lateness (Ignore This)
0 Points

Q8 Typed(Ignore this)
5 Points

Q9 PDF Submission
0 Points

No files uploaded

HW1 - Written Assignment

● Ungraded

Student

Gaurang Ruparelia

View or edit group

Total Points

- / 106 pts

Question 1 Problem Classification		10 pts
1.1	Firms	5 pts
1.2	Products	5 pts
Question 2 ML Applications		10 pts
2.1	Classification	5 pts
2.2	Regression	5 pts
Question 3 (no title)		15 pts
3.1	Target Variable	3 pts
3.2	Continuous or Discrete	4 pts
3.3	Predictor	4 pts
3.4	Linear?	4 pts
Question 4 Continued		30 pts
4.1	Squared Error	6 pts
4.2	Gradient Contribution	6 pts
4.3	Squared Error 2	6 pts
4.4	Gradient Contribution 2	6 pts
4.5	RSS	2 pts
4.6	RSS 2	4 pts

Question 5 (no title)		26 pts
5.1	Design Matrix	1 pt
5.2	Target Vector	1 pt
5.3	Closed Form Solution	2 pts
5.4	w's	4 pts
5.5	RSS	4 pts
5.6	TSS	4 pts
5.7	R^2	4 pts
5.8	Portion of Variance?	2 pts
5.9	Prediction	4 pts
Question 6 Different Alphas		10 pts
6.1	Gradient	4 pts
6.2	Gradient Descent	6 pts
Question 7 Lateness (Ignore This)		0 pts
Question 8 Typed(Ignore this)		5 pts
Question 9 PDF Submission		0 pts