**Machine Learning PS2: Regression**

Template Write-Up

**NOTES:**

* Due to the nature of Q1 & Q2 (it is difficult to write mathematical expressions in MS Word), it is perfectly fine to submit a physical (handwritten) copy of your answers, if that is what you prefer. Use of LaTeX (or, alternatively, a scan of your work, if you prefer to write it physically) is recommended to help fill in this template document.
* Note that Q1, part 1 has a typo:
  + ~~J(~~**~~w~~**~~) = J(w1, w2, w3)~~ 🡪 J(**w**) = J(w0, w1, w2)
* Note that Q1, part 2, you may wish to normalize the values given in the table before computing one pass of gradient descent.
* Note that for Q2B part (c), J(**w**) decreases with each pass of gradient descent, but will *never* actually converge to zero. Find a workaround – either implement a pass limit, ***or*** stop doing passes if the loss function J(**w**) isn’t changing much from pass to pass (define a threshold).

Q1-1. What is the cost function J(**w**) = J(w0, w1, w2) for the 5-item dataset in Problem 1?

|  |
| --- |
| Paper |

Q1-2. What are the values of (w0, w1, w2) after one iteration of gradient descent? Comment on whether you normalized the data or not for this pass.

|  |
| --- |
| Normalized data  Feature vector is[ -1.06297949e-16 ,8.54987593e-01 ,4.42261117e-01] |

Q2A-a. Given a set of numbers x1,…,xm, write down the equations for the mean and standard deviations of these numbers.

|  |
| --- |
| Paper |

Q2B-a. Write down the formula for the loss function J(**w**) using the sum of the squared errors. Be sure to include a 1/2m term. Is it different from your answer to Q1-1? Why or why not?

|  |
| --- |
| Paper |

Q2B-b. Derive the partial derivatives of J(**w**) with respect to w0, w1, and w2.

|  |
| --- |
| Paper |

Q2B-d. For learning rates α = 0.1 and 0.4. Comment on which of the three learning rates gives the best result.

|  |
| --- |
| 0.4 -> Algorithm converges in after 92 steps  for 0.3 it happens after 125 and for in 0.1 in 365 |

Q2C. Use the **w** obtained in Q2B to predict the housing prices. Predict the price of a house with 1650 sq. ft. area and 3 bedrooms. Don’t forget to normalize the features when you make this prediction!

|  |
| --- |
| 293083 |

Q3. Any ideas on how to work with the YikYak data? Explain your approach and come up with a function f(xE,xS,xW,xN) which predicts the longitude and latitude of any YikYak message. Calculate and report the mean squared error across the data points.

|  |
| --- |
| Sorry, didn’t have time to tackle this one. |