Tasks regarding Theory of Gradient Descent and Normal Equation

Page 19

Solve the following tasks:

1. **Explain in your own words how gradient descent and the normal equation method work. What are they trying to minimize?**
   ANSWER:
   - Gradient Descent iteratively adjusts the model parameters to minimize the cost function by "descending" in the direction of the steepest gradient.
   - The Normal Equation directly computes the optimal values of the parameters that minimize the cost function by solving a matrix equation.
   - Both methods aim to minimize the same cost function, which quantifies how well the model fits the data.
2. **Discuss the computational complexity of both gradient descent and the normal equation method. Which one do you think would be more efficient for very large datasets?**
   ANSWER:
   - Gradient Descent's computational complexity mainly depends on the number of iterations and the number of features. It is generally more scalable for larger datasets.
   - The Normal Equation involves matrix operations that can be computationally expensive ($O(n^3)$ for matrix inversion) when the feature set is large.
3. **Do both methods lead to the same values of *θ*? Run the code for both methods and compare the *θ* values. What do you observe?**
   ANSWER: Both methods aim to find the values of θ that minimize the cost function. In the case of this dataset we get:
   GRADIENT DESCENT: $\theta_0 = 1697.1$ and $\theta_1 = 29.33$
   NORMAL EQUATION: $\theta_0 = 1697.1$ and $\theta_1 = 29.33$
   With this dataset we get the same answer.
4. **Given the dataset we've worked on, which method do you think is more suited and why? Consider factors like the size of the dataset and the presence (or absence) of outliers.**
   ANSWER:
   - For smaller datasets with fewer features, the Normal Equation is often more suitable due to its direct computation.
   - For larger datasets or datasets with many features, gradient descent is usually more efficient.
5. **Imagine you'd have to find the global minima of the function below. Which method of the above would you choose and why?**
   ANSWER:
   Normal Equation. Gradient descent might get stuck at a local minimum, depending on the previously chosen starting point.

Page 23:

1. Do you observe a correlation between the input and output parameters?
   Currently the data set is too small (5 Points); therefore, this answer could/should be adjusted in the future!
   ANSWER: Possible explanations would be a linear or quadradic relation between speed (laser power) and the area of the weld seam.

2. Considering the equation defining laser energy density, explain why do some points have the same speed or the same power, whereas the corresponding area of the weld seam varies?
   ANSWER: In the dataset, there are tests with the same speed. In these cases, the power varies leading to variations in the area of the weld seam. Similarly, there are points with similar power, where the speed varies, resulting in variations in the area of the weld seam.

Page 28:

1. What can be understood from this diagram?
   Currently the dataset is too small (5 Points); therefore, this answer could/should be adjusted in the future!
   ANSWER: The diagram illustrates the results of the gradient descent. According to the diagram, increasing the power with a constant speed scarcely affects the area of the weld seam. Nevertheless, increasing the power should logically leads to a larger weld seam.
   On the other hand, increasing the speed with a constant power markedly reduces the area of the weld seam. The predominantly negative values for the area of the weld seam raise concerns about the accuracy of the model.
   Since all the 5 points are above the hypothesis, gradient descent could not provide a reasonable fit for the parameters.

2. Is the hypothesis function as expected? Justify your answer!
   Currently the data set is too small (5 Points); therefore, this answer could/should be adjusted in the future!
   ANSWER: No, the value of the hypothesis is below all datasets, suggesting that the algorithm did not find the global optima and rather a local optimum.

Page 31:

1. Find the optimum hypothesis predicting <u>weld strengths</u> using the same data set.
   ANSWER: In the MATLAB code change the line 224 as following:

```
y = T.Weld_Seam_Area; -> y = T.Strength;
```

The same applies for normal equation (line 276):

```
y = T.Weld_Seam_Area; -> y = T.Strength;
```

2. In the case where the thickness of the transparent substrate should be added to the prediction. Where has the script to be changed? Plots can be used for visualization purposes to understand the solution.
   ANSWER: In the MATLAB add following changes:
   Line 222:
   x3 = T.TH;
   X = [x1, x2, x3];

   Line 275:
   X=[x1, x2, x3];

3. A lap shear specimen with a new sets of input parameters has been welded and its tensile strength has been evaluated. Does the result of the test match the prediction of the previously identified hypothesis function h? Add this new input at the end of the "Training_Data.xlsx". This will help to improve the quality of the prediction.

   This task can only been completed after measuring new data in the lab, and including them in the database used to run ML algorithms