NYCU Introduction to Machine Learning, Homework 1

[111550149], [林悅揚]

Part. 1, Coding (60%):

(10%) Linear Regression Model - Closed-form Solution

1. (10%) Show the weights and intercepts of your linear model.

```
2024-10-03 16:09:58.360 | INFO     | __main__:main:79 - LR_CF.weights=array([2.8491883 , 1.0188675 , 0.48562739, 0.1937254 ]), LR_CF.
intercept=-33.8223
```

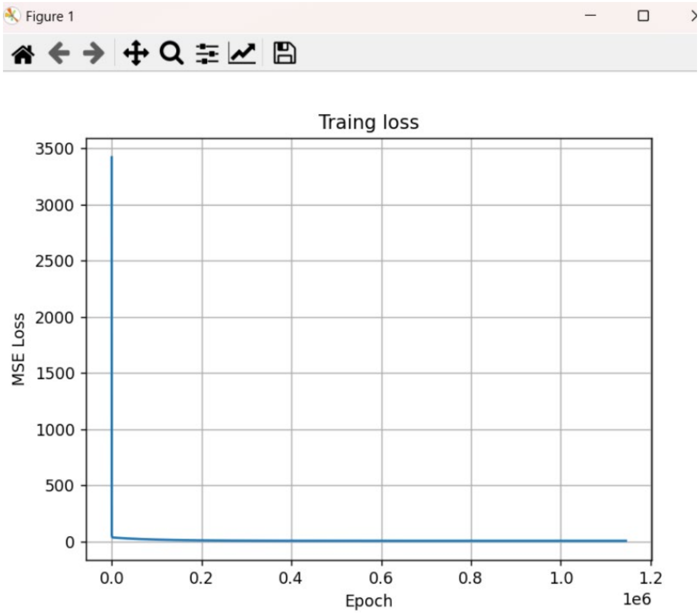
(40%) Linear Regression Model - Gradient Descent Solution

2. (10%)
   - Show the hyperparameters of your setting (e.g., learning rate, number of epochs, batch size, etc.).
   - Show the weights and intercepts of your linear model.

```
losses = LR_GD.fit(train_x, train_y, learning_rate=1e-4, epochs=1145000)
```

```
2024-10-03 16:12:54.728 | INFO     | __main__:main:84 - LR_GD.weights=array([2.84758402, 1.01836401, 0.48060839, 0.19257529]), LR_GD.
intercept=-33.7390
```

3. (10%) Plot the learning curve. (x-axis=epoch, y-axis=training loss)



4. (20%) Show your MSE.cf, MSE.gd, and error rate between your closed-form solution and the gradient descent solution.

```
2024-10-02 17:49:14.976 | INFO     | __main__:main:118 - Mean prediction difference: 0.0111
2024-10-02 17:49:14.977 | INFO     | __main__:main:123 - mse_cf=4.1997, mse_gd=4.1986. Difference: 0.027%
```

(10%) Code Check and Verification

5. (10%) Lint the code and show the PyTest results.

```
=========================== test session starts ============================
platform win32 -- Python 3.9.19, pytest-7.4.4, pluggy-1.0.0
rootdir: C:\Users\USER\Downloads\release
collected 2 items

test_main.py 2024-10-03 16:44:48.806 | INFO     | test_main:test_regression_cf:27 - model.weights=array([[3.]]), model.intercept=array([4.])
.2024-10-03 16:44:49.574 | INFO     | test_main:test_regression_gd:39 - model.weights=array([3.]), model.intercept=3.9999966785390386
.

============================ 2 passed in 1.44s =============================
```

Part. 2, Questions (40%):

1. (10%) How does the presence of outliers affect the performance of a linear regression model? How should outliers be handled? List at least two methods.
   Outliers are extreme data points. They may pull the regression line toward them, which can result in biased predictions. Outliers increase the overall error of the model, leading to poor fitting for the majority of the data and reduced model accuracy.
   Handling outilers:
   Method 1: Removing Outliers: If removing the outliers does not significantly reduce the sample size or introduce bias, this approach can be effective. Removing outliers helps eliminate their influence on the regression line, improving the accuracy of the model.
   Method 2: Replacing Outliers with Reasonable Values: If the outliers are not errors but extreme values, they can be replaced with more reasonable values. This can be done by capping the values at a certain threshold. This method helps maintain the integrity of the dataset while minimizing the impact of outliers on the regression model.
2. (15%) How do different values of learning rate (too large, too small…) affect the convergence of optimization? Please explain in detail.

   Too large learning rate: If the learning rate is too large, the model may overshoot the optimal point, causing it to diverge or fail to converge, leading to instability in the parameter updates.
   Too small learning rate: If the learning rate is too small, convergence becomes very slow, and it may take a long time for the model to reach the optimal solution, or the model may get stuck in a local minimum.

3. (15%)
   - What is the prior, likelihood, and posterior in Bayesian linear regression. [Explain the concept in detail rather than writing out the mathematical formula.]
   - What is the difference between Maximum Likelihood Estimation (MLE) and Maximum A Posteriori Estimation (MAP)? (Analyze the assumptions and the results.)

   Likelihood refers to the probability density values of the residuals given certain parameters. In other words, it is the probability of the observed data given a specific set of parameters. This plays a central role in estimating the parameters after observing the data.

Prior refers to the probability density values of the parameters before observing the data. It reflects our beliefs or assumptions about the parameters prior to any data observation, which could be based on previous research or expert knowledge.

Posterior refers to the probability density values of the parameters given the residuals. It combines prior knowledge and the likelihood, representing our updated understanding of the parameters after incorporating the observed data.

Maximum Likelihood Estimation (MLE):

Assumption:
MLE assumes that the parameters of a statistical model are fixed but unknown, and we need to estimate these parameters based solely on the observed data. It does not incorporate any prior beliefs or knowledge about the parameters.
Result:
MLE often works well when there is a large amount of data. However, it can lead to overfitting in cases where the dataset is small, as it attempts to fit the model as closely as possible to the data, making the estimates sensitive to random fluctuations or noise in the data.
Maximum A Posteriori Estimation (MAP):

Assumption:
MAP assumes that the parameters are random variables with a known prior distribution, which expresses any prior beliefs about the parameters before observing the data. This prior is combined with the observed data (likelihood) to make parameter estimates.
Result:
MAP can reduce overfitting compared to MLE, particularly when the dataset is small. By introducing regularization (bias from the prior), it reduces the variance in the parameter estimates. If the prior is weak or uniform, MAP approaches MLE. However, if the prior is strong, the results will be more biased toward the prior.