

Lab Course Machine Learning

Exercise Sheet 2

Prof. Dr. Dr. Lars Schmidt-Thieme

Jung Min Choi

HiWi: Harish Malik

Submission deadline : October 31, 2024

1 Linear Regression on Time Series Data (12 points)

1. **[2 points]** Time Series Exploration: Try to understand the *time series* dataset uploaded along with the exercise. This involves (but is not limited to) plotting the multivariate time series (time on the x-axis), clearly labeled and formatted, understanding how the multivariate time series interacts, exploring the correlation between the different variables, plotting variable density functions, identifying inherent seasonality or trend, etc.
2. **[1 points]** Train/Test split: As a next step, split the data into train, validation, and test by considering the following rules:
 - a) Train end: 2017-06-26
 - b) Validation end: 2017-10-24
 - c) Test end: 2018-02-21

Draw a single plot of the entire time series (for the target variable) and clearly mark within the plot the train end, validation end, and test end.

From now on, you are working with the train, validation, and test datasets.

3. **[1 points]** Scaling data: For each of these datasets, perform standard scaling to scale each variable independently, i.e.,

$$z = \frac{(x - \text{mean}(x))}{\text{std}(x)}$$

Check if the standard scaling worked by computing the mean and standard deviation for a particular variable within the data.

4. **[6 points]** Regression: Implement linear regression using the Least Squares Estimate (LSE) method. You are required to:
 - a) Implement matrix multiplication and inversion using `numpy`.
 - b) Compute the estimated parameters using the LSE method.
5. **[2 points]** Loss: Using the parameters ($\hat{\beta}$) that were learned, calculate the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) for the train, validation, and test splits of the dataset. Additionally, plot the true target and predicted target values for the train, validation, and test splits on the same plot.

2 Visualizing Least Squares Estimation on a Toy Dataset (8 points)

1. **[1 point]** Toy Dataset: Load the toy dataset from the provided file `toy_data.csv`. You do not need to do preprocessing.
2. **[4 points]** Linear Regression using Least Squares: Implement the linear regression model using the Least Squares Estimation (LSE) method. Given a matrix of input features X and a vector of target values y , compute the parameters $\hat{\beta}$ using `numpy` (pure Python). You should not use any built-in functions for linear regression. Instead, derive the result based on the least squares criterion.
3. **[3 points]** Visualization:
 - a) Visualize the linear regression fit by plotting the toy dataset and the predicted regression plane in a 3D plot using `matplotlib`.
 - b) The x-axis and y-axis should represent the two input features, and the z-axis should represent the target variable.
 - c) Plot the ground truth data points and the estimated regression plane.
 - d) Visualize the residuals by drawing lines connecting each ground truth point to the corresponding point on the regression plane, representing the difference between actual and predicted values.
 - e) Use different colors to represent the actual data points, the regression plane, and the residuals. Ensure the plot is clearly labeled and formatted for better readability.