

Assignment, Week 4

Due date: 10 Feb 2024

1. The closed-form solution for linear regression for multiple independent variables is:
 $(X^T X)^{-1} X^T y$
Derive the expressions for slope and intercept for the special case of $y = mx + c$.
2. Assignment from notebook [Lec4 LinReg from scratch.ipynb](#):
 - a. Increase the number of epochs (`n_epoch`) to 50 in the stochastic gradient descent optimization method. Plot the resulting Θ vs. `n_epoch`.
 - b. Repeat the same for the mini-batch gradient descent method.
 - c. How does the evolution in Θ as a function of `n_epoch` differ between the two methods above
3. Assignment from notebook [HO-4-1-house_price-3_include_categorical_features.ipynb](#):
 - a. Correct some of the mistakes in the notebook
 - b. Use of imputers other than `SimpleImputer`
 - i. `KNNImputer`
 - ii. `IterativeImputer`
 - c. How does transforming one of the features which have a long tail distribution (eg. `total_bedrooms`) to a bell curve impact RMSE (train and test)
4. Assignment from notebook [HO-4-2-house_price-4_pipeline.ipynb](#): How does not performing feature scaling affect the resulting regression fit?
5. Assignment from notebook https://github.com/raHO-4-5-polynomial_regression_sklearn.ipynb: Find the optimal regularization parameters (λ) for (a) Ridge Regression and (b) Lasso Regression. Make sure to cover a wide range of λ for this exercise.
6. Go through every Google Colab notebook for [week 4](#) and refer to the suggested resources. No need to submit any answer for this.