3.1 Boston House-Prices

- Q: *Which one gets better results? If you were not able to implement one of them, which one you suspect should get better results and why?*
  A: KNN gets better result because linear regression underfits the data
- Q: *Which measures are you using to compare them?*
  A: Mean Squared Error
- Q: *Which one is faster during training? And during inference time?*
  A: KNN is faster during traning. Linear regression is faster during inference.

3.2 Breast Cancer Wisconsin

- Q: *Which one gets better results? If you were not able to implement one of them, which one you suspect should get better results and why?*
  A: Logistic regression gets better results, because it builds precise decision boundary and KNN is confused a lot in that zone.
- Q: *Which measures are you using to compare them? Which algorithm has more False Positives? And False Negatives? If you were not able to implement one of them, explain the concept of False Positives and False Negatives.*
  A: I used F1 score and confusion maxtix. Logistic regression performed better. It had 1 False Positive and 1 False Negative vs 5 and 2 for KNN.
- Q: *If you were a doctor, which algorithm would you use for the prediction? If you were not able to implement one of them, explain how you would use a confusion matrix to answer this question.*
  A: I would use the algoritm with the least False Negative results. In my case that was Logistic Regression.

4 Using Scikit-learn and comparing results

- Q: *Which one gets better results? Please specify per dataset.*
  A: My Linear regression performed a little bit better, because by default it doesn't have regularization and better fits the data. But it was slower as I implemented batch gradient descent which is less efficient in terms of learning time.
  Sci-kit's Logistic regression performed better, because my gradient decent had hard time finding precise coefficients.
- Q: *How could you modify the other methods to achieve similar results?* A: For linear regression I need to enable regularization and for logistic regression I need to change algorithm for finding more precise coefficients.
- Q: *Which one is faster during training? And during inference time?*
  A: Sci-kit learn is faster during traning and they are identical during inference.
- Q: *Using the methods from Scikit-Learn, play around with the parameters. What is the best result you can get? What parameter was the most important for each dataset?*
  A: For both dataset the most important was max_iterations and learning_rate. Penalty and alpha also influence the score, but only in a negative way.
- Q: *(Bonus) For Linear Regression, plot the coefficients after training models with L1 and L2 regularization. How do they differ?*
  A: L1 removes unimportant features, L2 makes coefficients smaller.
- Q: *(Bonus) For Logistic Regression, plot the coefficients after training models with L1 and L2 regularization. How do they differ?*
  A: L1 removes unimportant features, L2 makes coefficients smaller.

5 How long did it take?

40 hours