HDS 5230 - Instructions for application assignment - week 11

For this week's assignment, you are required to investigate the accuracy-computation time tradeoffs of the different optimization algorithms (*solvers*) that are available for fitting linear regression models in Scikit-Learn. Using the code shared via the Python notebook (part of this week's uploads archive) where the use of logistic regression was demonstrated, complete the following operations:

1. Among the different classification models included in the Python notebook, which model had the best overall performance? Support your response by referencing appropriate evidence.

2. Next, fit a series of logistic regression models, without regularization. Each model should use the same set of predictors (all of the relevant predictors in the dataset) and should use the entire dataset, rather than a fraction of it. Use a randomly chosen 80% proportion of observations for training and the remaining for checking the generalizable performance (i.e., performance on the holdout subset). Be sure to ensure that the training and holdout subsets are identical across all models. Each model should choose a different solver.

3. Compare the results of the models in terms of their accuracy (use this as the performance metric to assess generalizability error on the holdout subset) and the time taken (use appropriate timing function). Summarize your results via a table with the following structure:

Solver used	Training subset accuracy	Holdout subset accuracy	Time taken

4. Based on the results, which solver yielded the best results? Explain the basis for ranking the models - did you use training subset accuracy? Holdout subset accuracy? Time of execution? All three? Some combination of the three?

Submit your work in the form of a Python notebook that contains all of the appropriate (and only the appropriate) code needed to address the questions, along with accompanying explanations to answer the above-listed questions. Provide appropriate comments in your code to annotate it, highlighting key decisions you have made to address the questions. In addition, include your name and assignment number at the beginning of the notebook.