

Optimization for Machine Learning (Spring 2026): Homework 3

- LO4ML stands for the textbook *Linear Algebra and Optimization for Machine Learning*.
 - You **must email your submission** to kbala@wsu.edu. The main document (report) should be in **PDF format** and its name should identify you in this manner. If you are Billy Circlovich, you should name your submission BillyCirclovich.Hw3.pdf. **Please avoid white spaces in the file name.**
 - To submit your code/implementation, you should **include all files in a compressed folder**, e.g., zipped or tar-gzipped, named in the same manner, i.e., **BillyCirclovich.Hw3.zip**. You could include the PDF file inside this zipped folder, along with the other files.
 - **WSU Email will not allow attachments of .py files, even inside Zipped folders! You can save such files as .txt files, or include corresponding Jupyter notebooks (.ipynb files) instead.**
 - **Begin the SUBJECT of your email submission with the same FirstnameLastname, expression, e.g., “BillyCirclovich Hw3 submission”.**
 - **This homework is due by 10:00 PM on Tuesday, March 10.**
1. (10) LO4ML Problem 5.5.2 from Page 225. For this as well as other problems, you are encouraged to use the notation we have seen in class: w for the variable (parameter) vector and x_i for the data vector of the i -th instance, etc., rather than the notation used by LO4ML, e.g., W, X_i , etc.
 2. (10) LO4ML Problem 5.5.4 from Page 228. Use the notation of P_w and U_w (seen in class) in place of Λ_w^u and Δ_w^p used in LO4ML.
 3. Implement from scratch (using only standard vector/matrix operations and functionalities using, e.g., numpy) the following optimization routines for L_2 -SVM and logistic regression (LR) SVM, both with Tikhonov regularization:
 - (a) (20) basic gradient descent with fixed learning rate (GD);
 - (b) (30) Adam variant of gradient descent (Adam); and
 - (c) (40) Newton method with line search (Newton).
 4. (60) You should test your implementations on standard binary classification datasets from, e.g., the UCI ML Repository (or from sklearn.datasets in Python). Two options are the Breast Cancer Wisconsin dataset ($n = 569, d = 30$) and the Adult Income census dataset ($n = 48842, d = 14$), but you can choose others. Make sure you test on at least one small, one medium, and one large dataset (as measured by n , the number of samples). Report how your implementations of the three methods (GD, Adam, Newton) perform by plotting the loss function values against number of iterations as well as against running times. Also compare the performance as measured by the final classification accuracy on a selected medium sized dataset for the two SVMs (L_2 vs LR).
 5. (30) Compare your Adam implementation with PyTorch (using torch.optim.Adam) for each dataset you study. Does your implementation give the same solution as PyTorch (up to numerical tolerances)? How do the running times compare, especially on the medium and large datasets?

Submit a brief report with your responses and output+analysis.