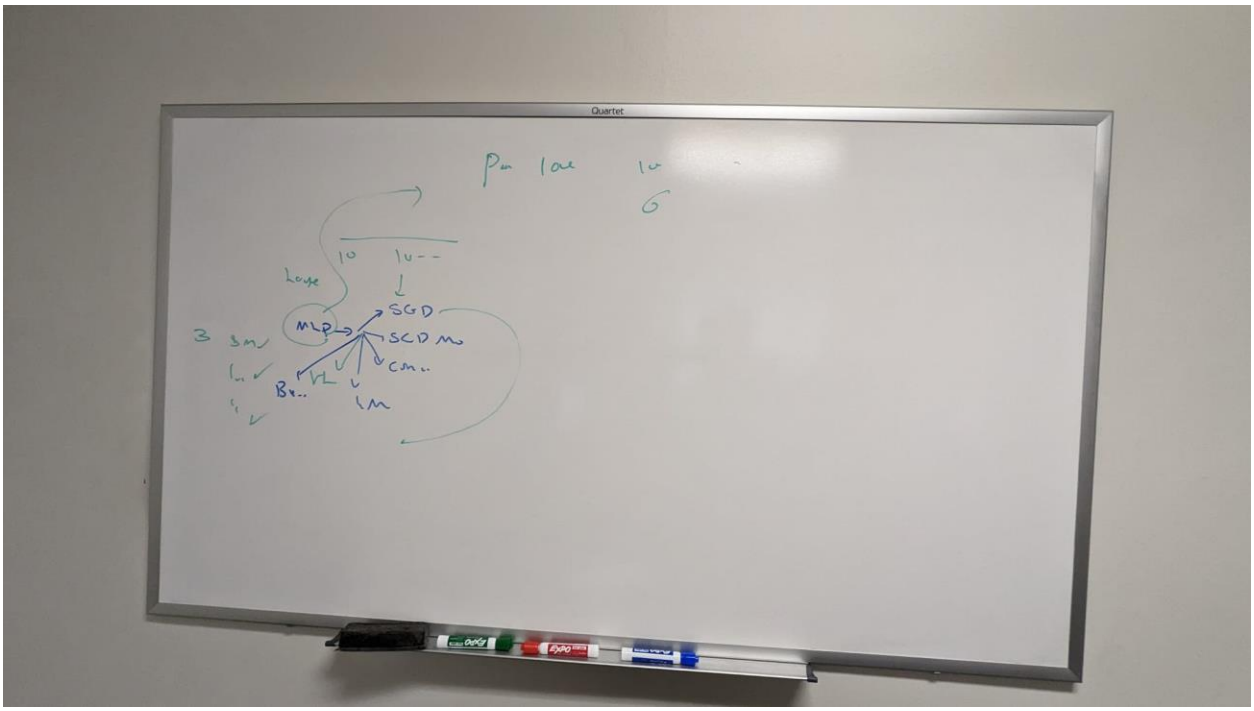

The objective of the project is – Compare optimization techniques learned in class – SGD, Variable Learning rate, SGD with momentum, Conjugate gradient, and LM. This should be a generalized code to have the number of layers as input (number of neurons), As general as possible. ++ Good to have will be to implement Bayesian regularization and/or early stopping as well for bonus content.

Question – Can we have a variable transfer function and still use these optimization techniques? Yes. Because architecture is no longer MLP then. Can we also have variable metrics (SSE, MSE, Absolute, etc.) – use one depending on problem statement – mse for regression, sse for classification.

Final comparison –

1. Varying dataset number of features
2. Varying number of layers – parameter numbers
3. Varying transfer functions
4. Varying Learning rate

Procedure - Pick 3 different datasets – Small, medium, and large (# of features) Experiment with each dataset, varying the number of layers to increase problem complexity and compare the performance of the optimization techniques.