Lab 1

A measure of model performance is called ‘loss’ when we are training/optimizing but is called ‘performance metric’ when testing on unseen data.

Most of this lab is a basic insight into generating data, fitting/training a linear model and evaluating it using a chosen metric.

In section 6, the concept of regularization is introduced. Regularization is done to prevent overfitting and improve the model's generalization ability.

The first method introduced is the L2 Regularization, or ridge regression. This term adds a penalty to the sum of the squared magnitudes of the coefficients.

L2 regularization is applied as a single value to the loss function, but this single value is derived from the sum of the squared magnitudes of *each individual parameter* (excluding the bias/intercept term, which is typically not regularized)

L2 universally has a dampening impact on all parameters, pointing them towards zero as it is "artificially" inflating the loss function.

But it also disproportionately targets larger parameters though the squaring of the magnitudes. This is important because it stops any model becoming too dependent on some larger parameters.

The derivative of the regularization term is $2\lambda\beta k$, hence, larger parameters are impacted more, as well as, contributing to the loss function more in the first place.

By preventing overfitting, L2 regularization helps the model capture the underlying patterns in the data rather than memorizing the training examples, thus improving its ability to make accurate predictions on new data.

The second method introduced is L1 Lasso which is a term to the loss function and is done to avoid overfitting.

L1 regularization adds a penalty based on the sum of the absolute values of the coefficients.

Feature selection (sparsity) is the main aspect of L1.

It drives features of less importance to zero.

This essentially removes parameters from the model making it more sparse. The result is a simpler model which is more efficient due to having less dimensions.

It is not differentiable at zero which is what causes parameters to be driven to 0. As opposed to L2 which whilst it shrinks, may not drive params to exactly zero.

The lab finishes by highlighting the importance of using a random seed to allow for repeatability.