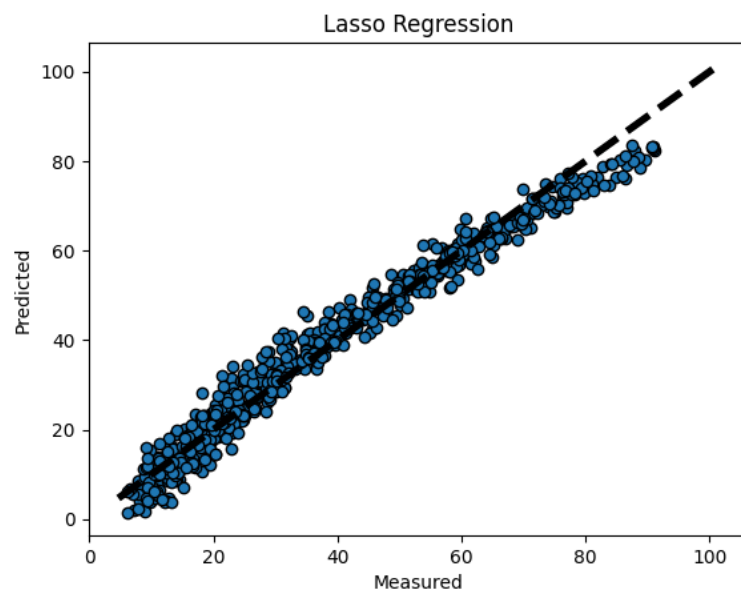


The dataset was split into train/test data sets and normalized to values between 0 and 1. A k-means algorithm was used to declutter the dataset into 3 different clusters.

1. Linear Regression

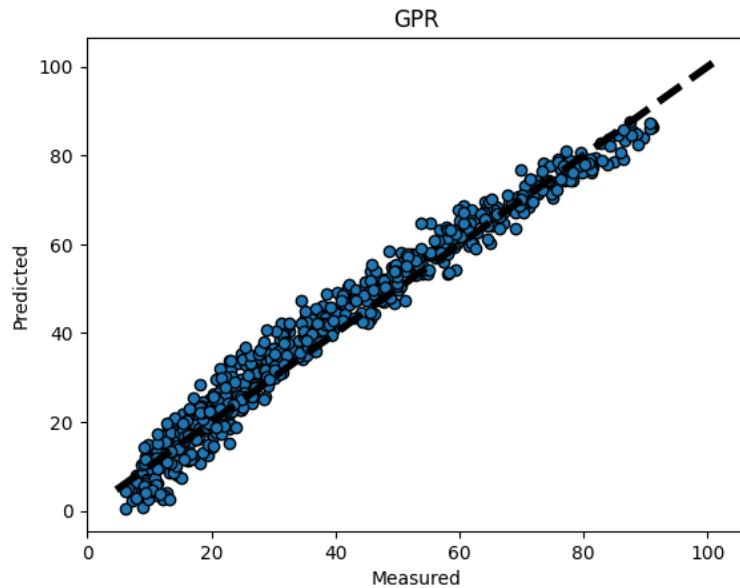
Lasso regression was implemented in this section. Lasso was chosen over Ridge because Lasso allows for some coefficients to be set to zero leading to a simpler model with a reduced set of coefficients. A repeated K-fold cross-validation was used to observe the model's performance. 10 folds were used and repeated 3 times. The mean MAE was 0.972 with a standard deviation of 0.004. The parity plot below shows the performance of the model predictions.



2. Gaussian Process Regression

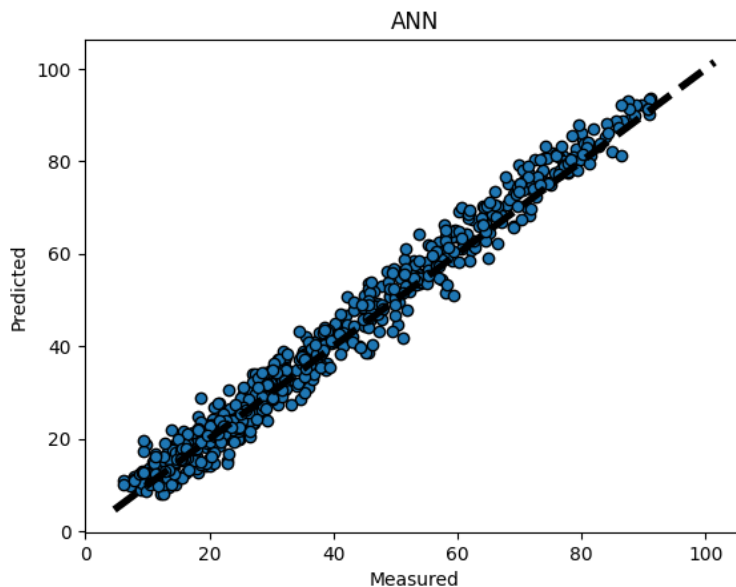
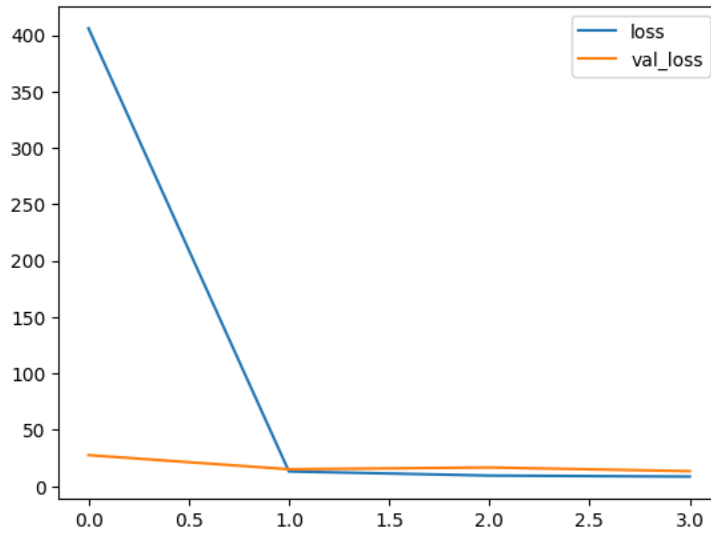
Not having any previous knowledge of the data and any possible trends, I used the anisotropic RBF kernel with a unique hyperparameter for each feature. Also, the built in hyperparameter optimizer was used to obtain the best values for each hyperparameter. This optimizer finds the optimum values by maximizing the log-marginal likelihood. K-fold cross

validation was used once again to determine the model performance. The MSE was found to be 4.975. The parity plot below shows the performance of the model predictions.



3. Neural Network

The network architecture had three hidden layers with 16 neurons each and used relu activation. Relu activation was chosen because it is the most widely used activation function. It is easy to compute and the derivative is easy to compute as well. The MSE was 3.66 and loss was minimized within 1 epoch. The parity plot below shows the performance of the model predictions.



All models were able to predict with fairly high accuracy. The neural network came out as the most accurate, but with only a slight improvement from GPR. Even though the model accuracy was marginal between GPR and ANN, ANN is much more efficient and faster. While the GPR model took over 10 minutes to complete, the ANN model only took about 10 seconds. The linear regression was also fairly efficient and compiled in under 30 seconds, but being a simple model, it has the possibility of not modeling the dataset correctly. Having some prior knowledge about

the dataset would have been beneficial to improve the implementation of the GPR method.

Having to blindly choose a kernel can impact the effectiveness of this method. ANN benefits from not having to have any knowledge of the behavior of the dataset and can model any complex behavior. But, caution must be taken to make sure that ANN does not create an over fit of the data.