Final Slide Deck Template

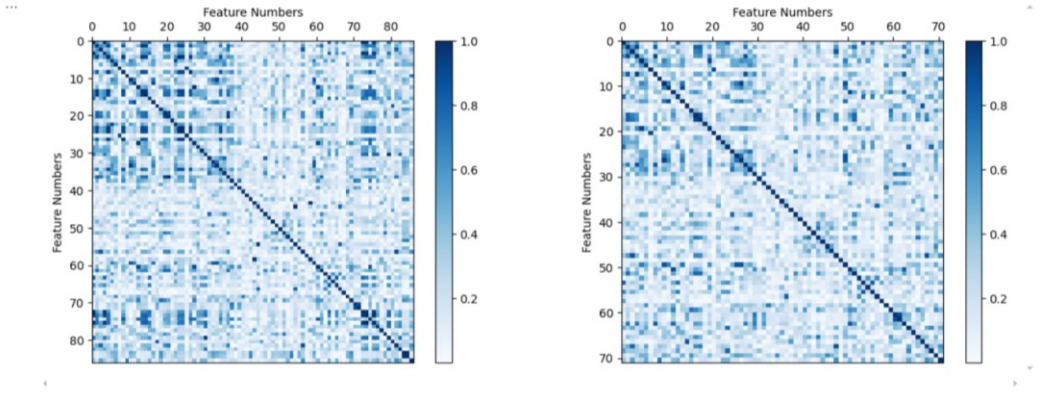
Jiahui Yang

Informatics Skunkworks (**MSE 401**), 3 Credits

08/11/2024

Assignment 2 – Complete through section 3 of module 1: Basics of machine learning

- Correlation matrix after removing highly correlated features above exercise 3.2




```
len(features_df_lowcorr.columns)
```

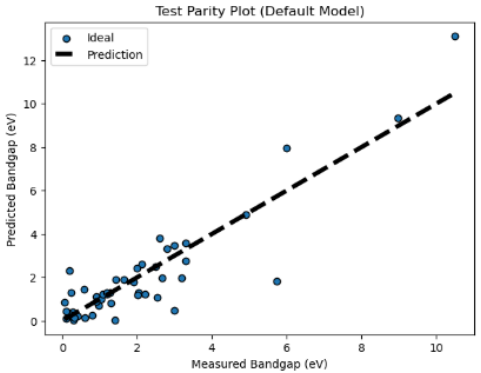
[48] ✓ 0.0s Python

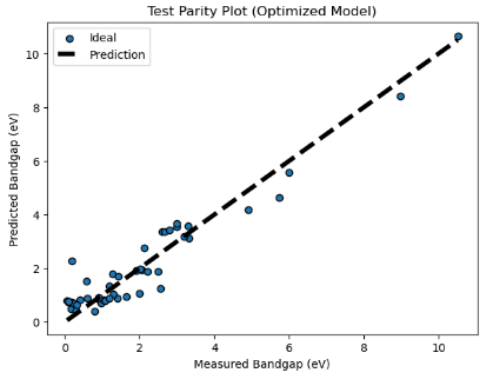
71

Emphasized the importance of data cleaning, removing highly correlated and constant features, and feature normalization to improve the performance and accuracy of machine learning models.

Assignment 3 – Complete through section 7 of module 1: Basics of machine learning

- Optimized model test data parity plot from exercise 6.4





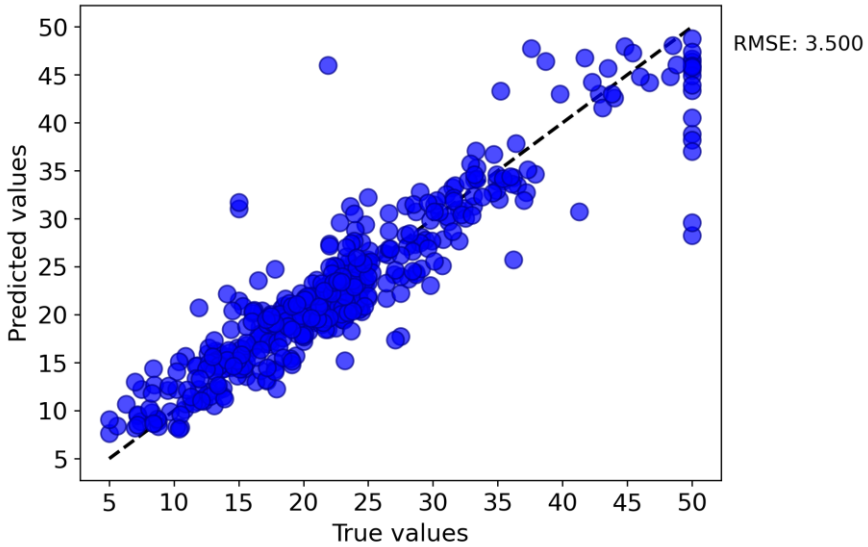
6]:

	Error Metric	Test Set (Default Model)	Test Set (Optimized Model)	Note
0	RMSE	1.0492 (eV)	0.6062 (eV)	(0.0 for perfect prediction)
1	RMSE/std	0.4884	0.2822	(0.0 for perfect prediction)
2	MAE	0.6811 (eV)	0.4776 (eV)	(0.0 for perfect prediction)
3	R2	0.7614	0.9204	(1.0 for perfect prediction)

By using cross-validation and a test set to evaluate model performance, we effectively avoided overfitting, ensuring the model's generalization ability and strong performance on unseen data.

Assignment 4 – Complete Introduction to MAST-ML activity

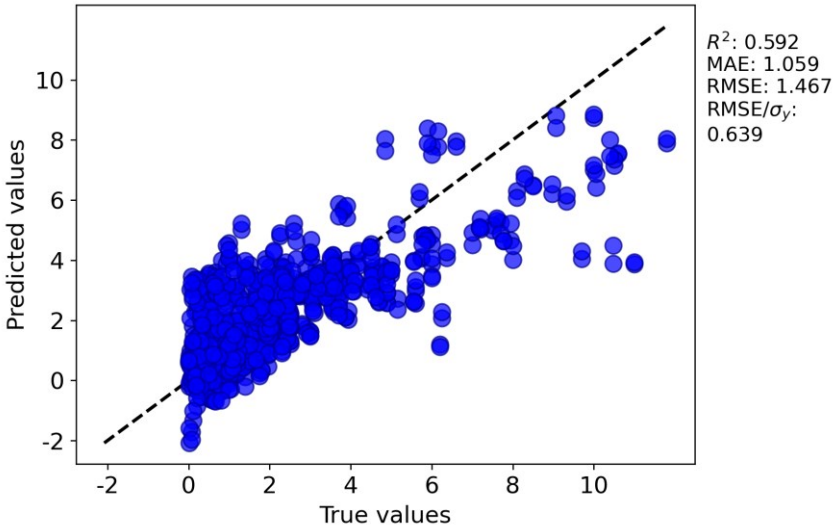
- Optimized model test data parity plot from exercise 6.4



Evaluating multiple machine learning models using cross-validation reveals the impact of model selection and parameter tuning on predictive accuracy and generalization.

Assignment 5 – Complete module 4: Comparing Model types

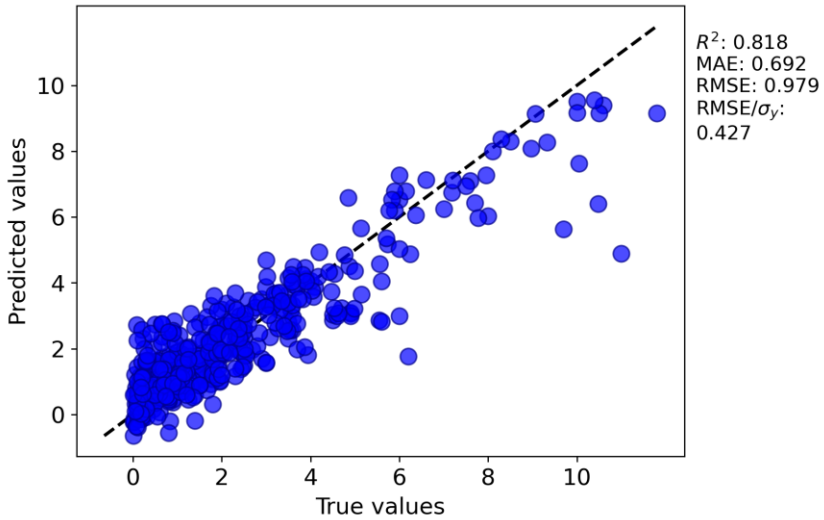
- Test data parity plot of your optimized model performance from Section 5 (5-fold cross-validation)



Using MAST-ML to reproduce machine learning workflows allows for automated generation of parity plots and statistical analyses, making it easier to evaluate and compare model performance, hyperparameter optimization, and cross-validation results.

Assignment 6 – Complete module 5: Hyperparameter Optimization

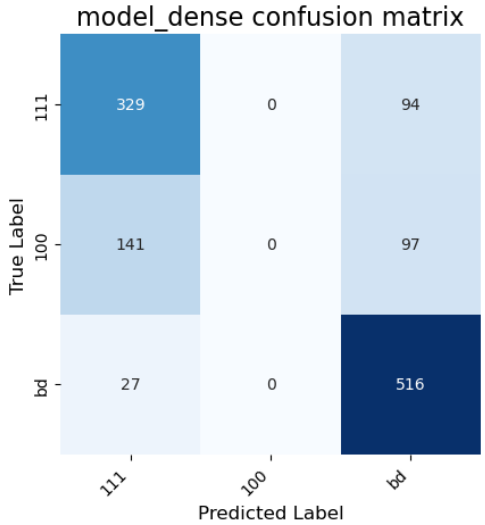
- Your parity plot of Test data for the best hyperparameters you found before changing the model neuron structure



Conducting iterative grid searches over multiple hyperparameters, including regularization, learning rate, and neural network structure, allows for a systematic and thorough optimization process, improving the model's performance by identifying the best combinations of these parameters.

Assignment 7 – Work Through the "ML4M - Image Data Activities.ipynb" notebook

- Your confusion matrix and F1 scores for Section 1 with the default Fully Connected Network (FCN)



Transfer learning with pretrained models significantly improves electron microscopy image classification accuracy, and data augmentation with dropout effectively mitigates overfitting.