SUPERCONDUCTORS

Ziping Wei

SUPERCONDUCTING MATERIALS

APPLICATIONS

- Magnetic Resonance Imaging (MRI)
- Superconducting Coils
- Frictionless Superconducting Wires

MAJOR ISSUES

- Superconductor conducts current with zero resistance only at or below its superconducting critical temperature (T_c)
- Predication of T_c is an open problem in the scientific community

Hamidieh, K. (2018). A data-driven statistical model for predicting the critical temperature of a superconductor. Computational Materials Science, 154, 346–354. doi: 10.1016/j.commatsci.2018.07.052

R CODES BY THE AUTHOR

https://github.com/khamidieh/predict_tc/blob/master/main_script_production_9.R

DATA SOURCE

UCI Machine Learning Repository
http://archive.ics.uci.edu/ml/datasets/Superconductivty+Data

Raw Data: 81 Features; 21,263 Superconductors with T_c value

- Atomic Mass
- First Ionization Energy
- Atomic Radius
- Density
- Electron Affinity
- Fusion Heat
- Thermal Conductivity
- Valence

⇨

- Mean
- Weighted mean
- Geometric mean
- Weighted geometric mean
- Entropy
- Weighted entropy
- Range
- Weighted range
- Standard deviation
- Weighted standard deviation

Number of Elements

80 features: properties of elements + 1 feature

Data used for the project: 80 Features; 1,000 Superconductors with T_c values

Features

80 standardized features
(dropped feature that describe number of elements of a superconductor)

Standardize Formula

$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}}$$

Target

Critical Temperature T_c

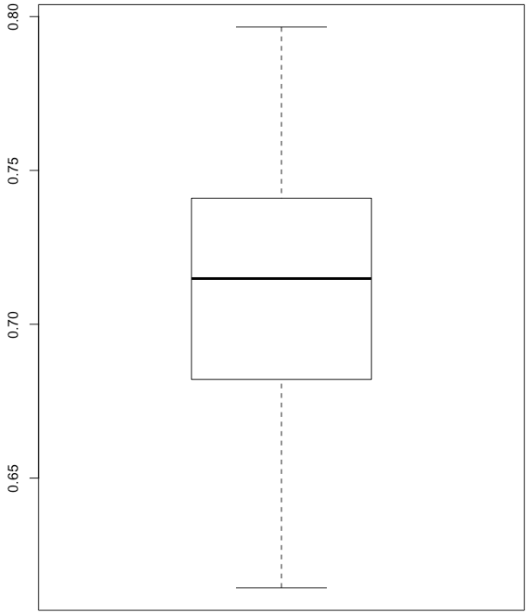
Observations

1,000 *(randomly selected from 21,263 observations)*

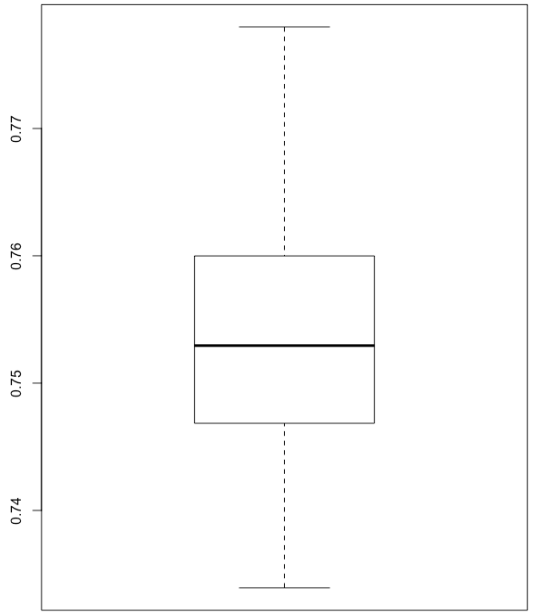
Models Applied

Lasso, Ridge, Elastic Net ($\alpha = 0.5$), Random Forest

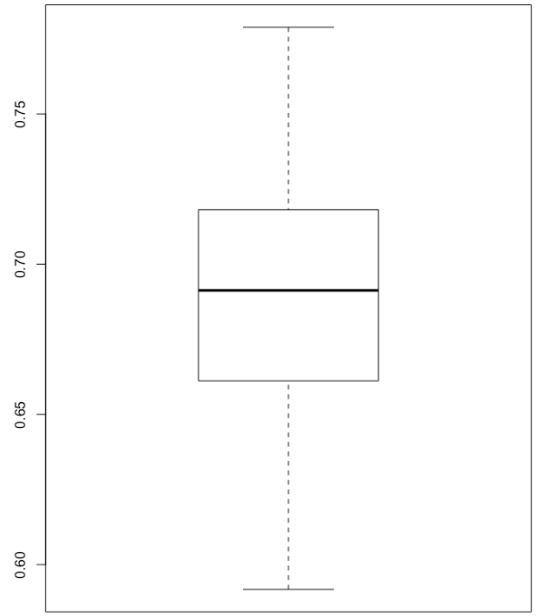
Boxplot of R^2_{test} & R^2_{train}



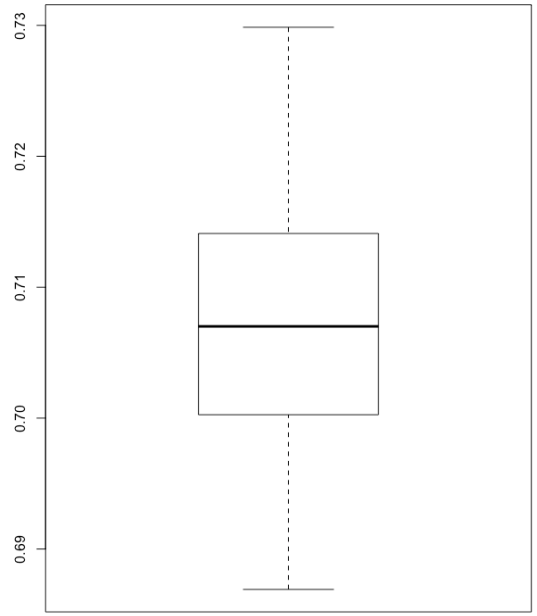
Lasso Test R_Square



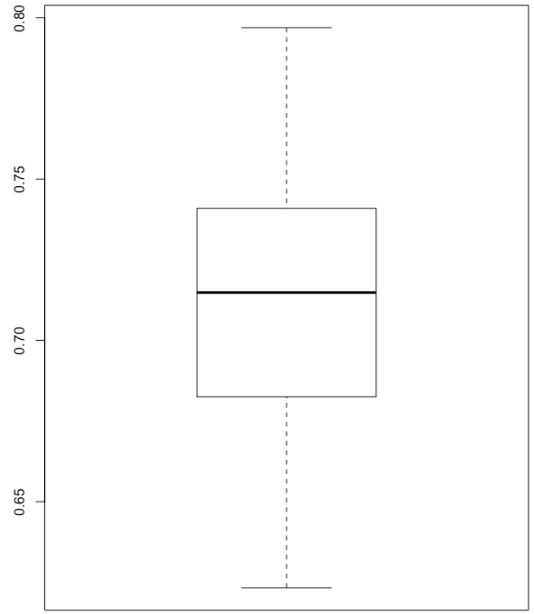
Lasso Train R_Square



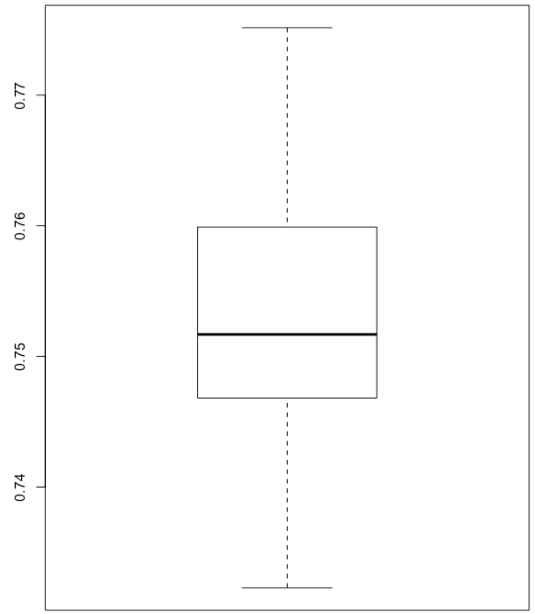
Ridge Test R_Square



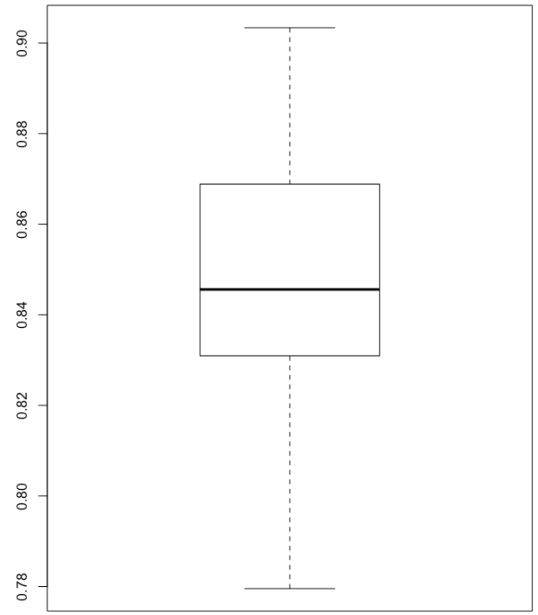
Ridge Train R_Square



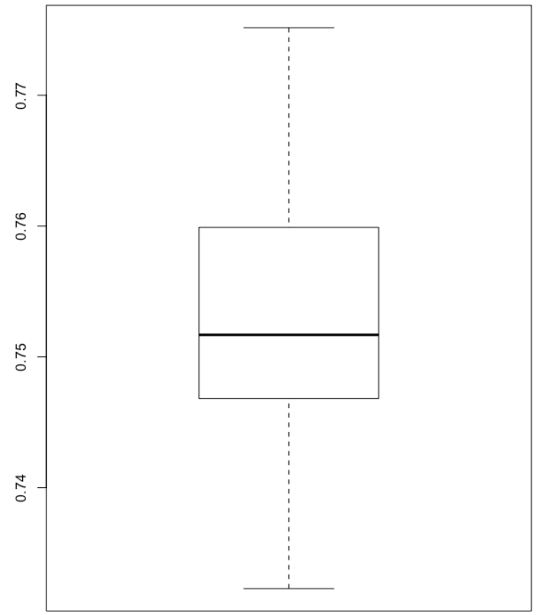
Elastic-Net Test R_Square



Elastic-Net Train R_Square

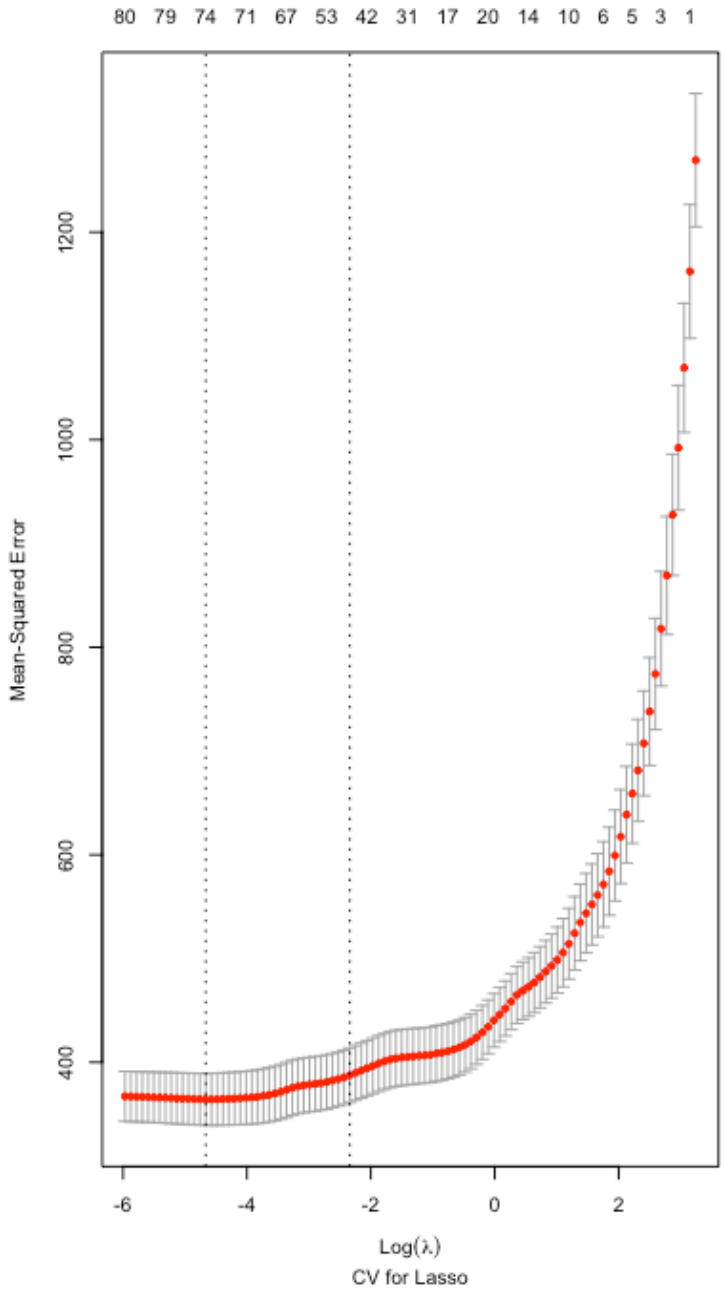


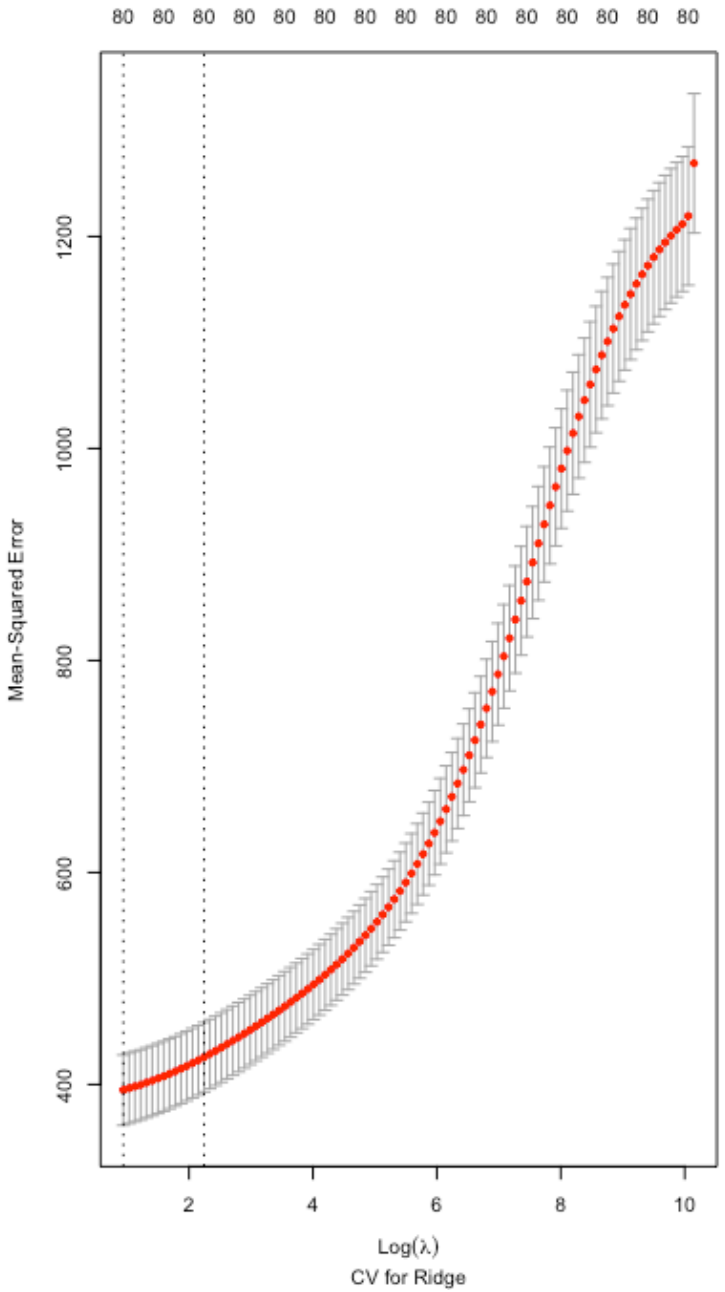
Random Forest Test R_Square

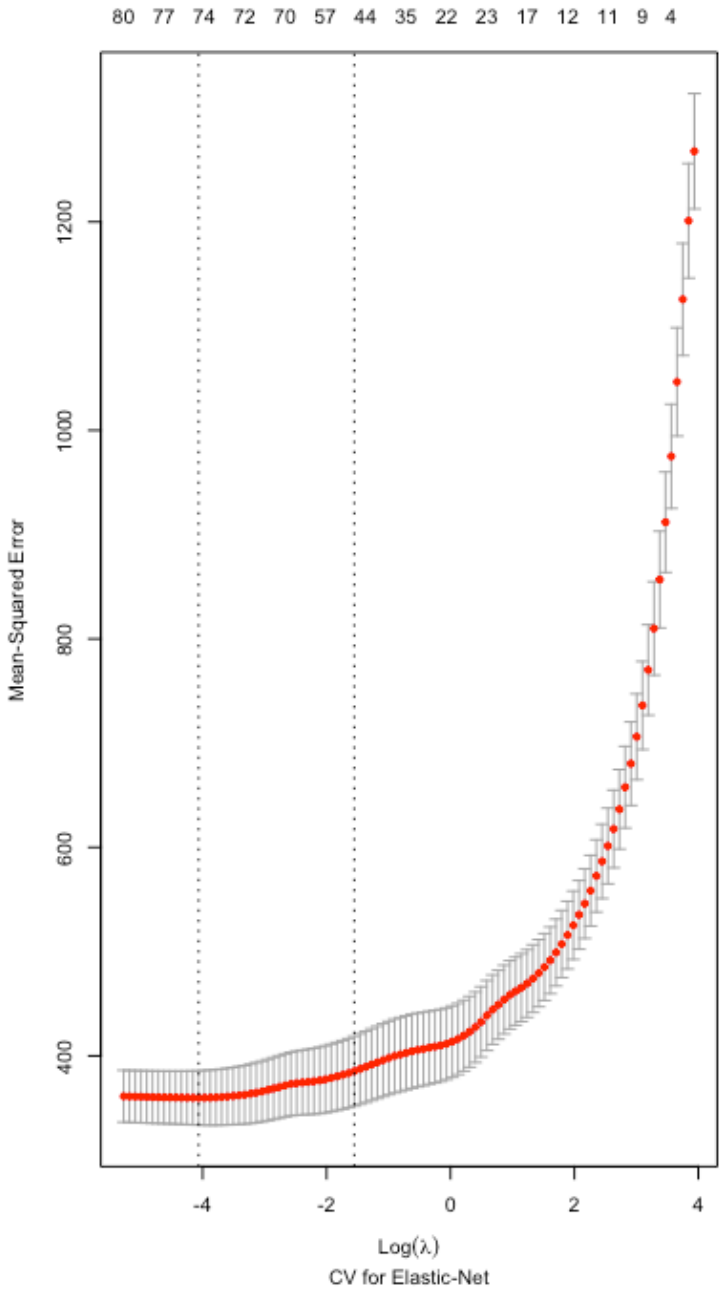


Random Forest Train R_Square

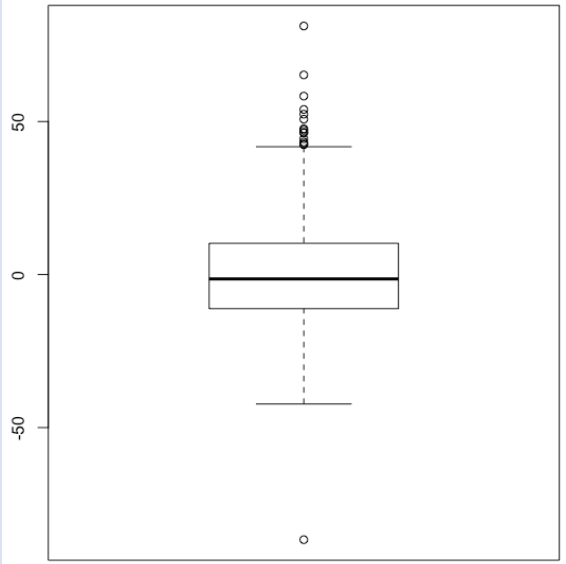
10-fold CV Curves for Lasso, Ridge and Elastic-net



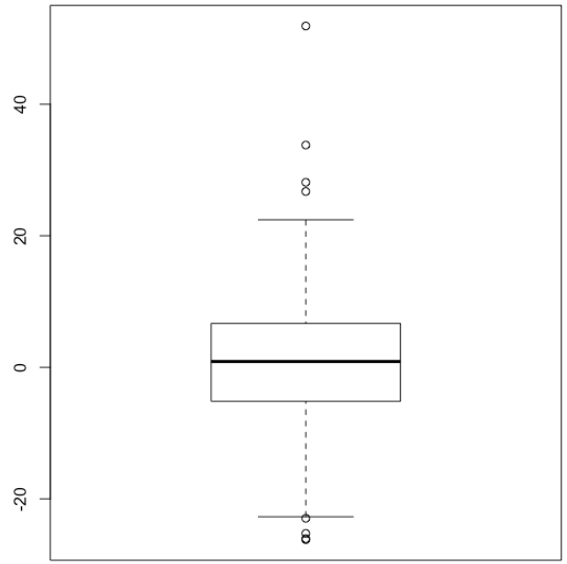




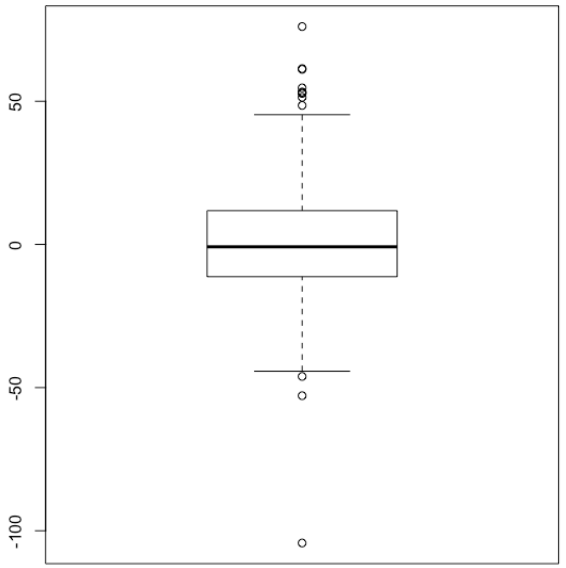
Boxplots of Train and Test Residuals



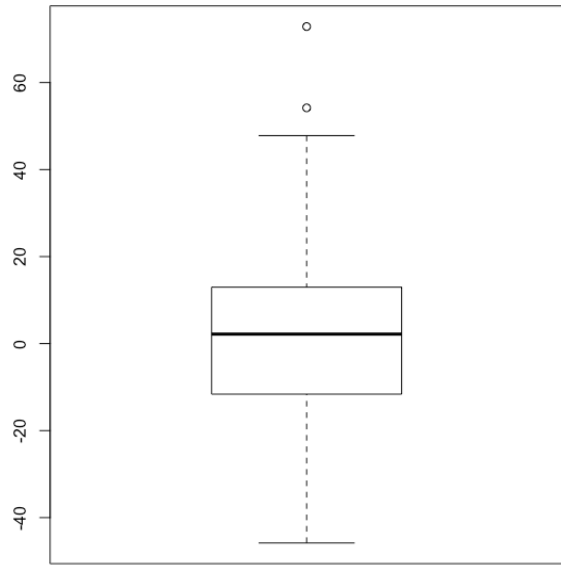
Residuals of Random Forest Train



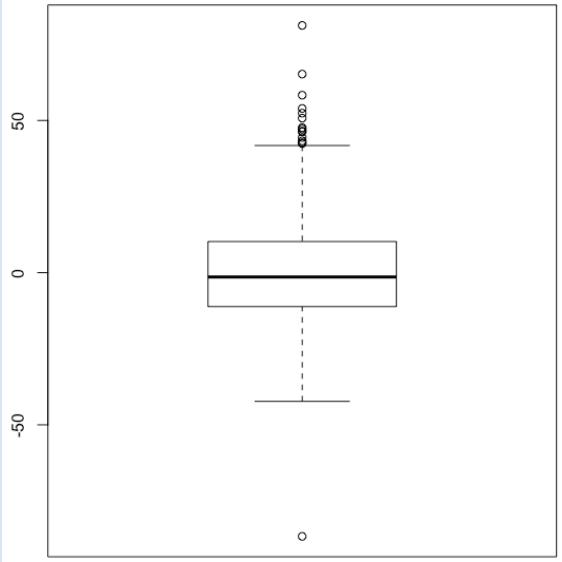
Residuals of Random Forest Test



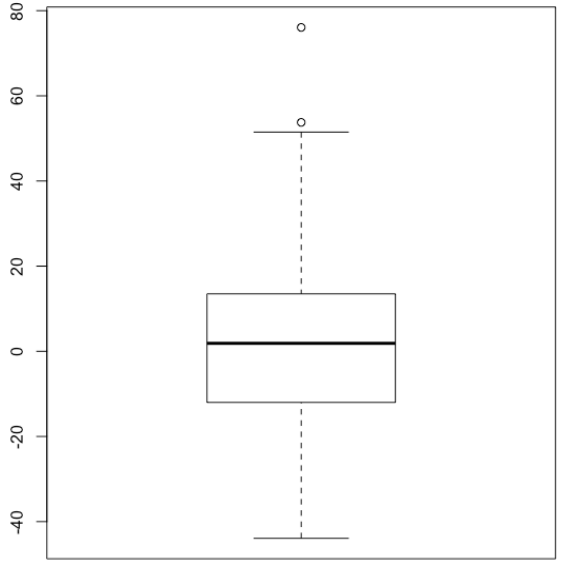
Residuals of Ridge Train



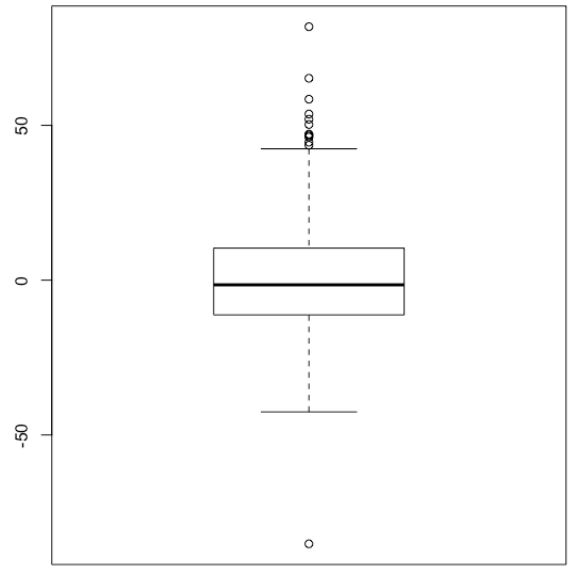
Residuals of Ridge Test



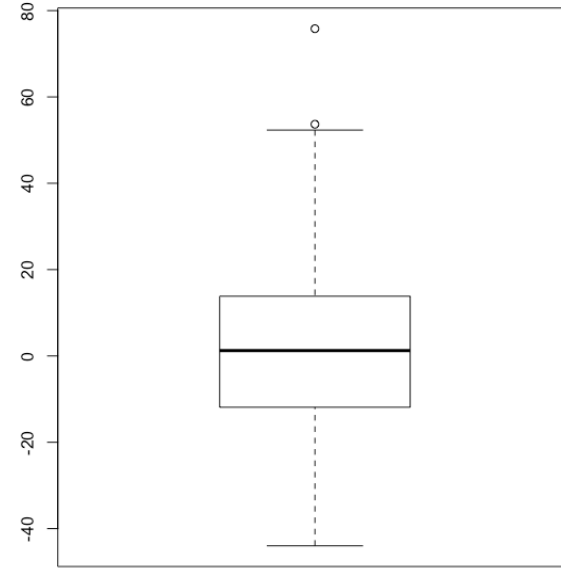
Residuals of Elastic-Net Train



Residuals of Elastic-Net Test

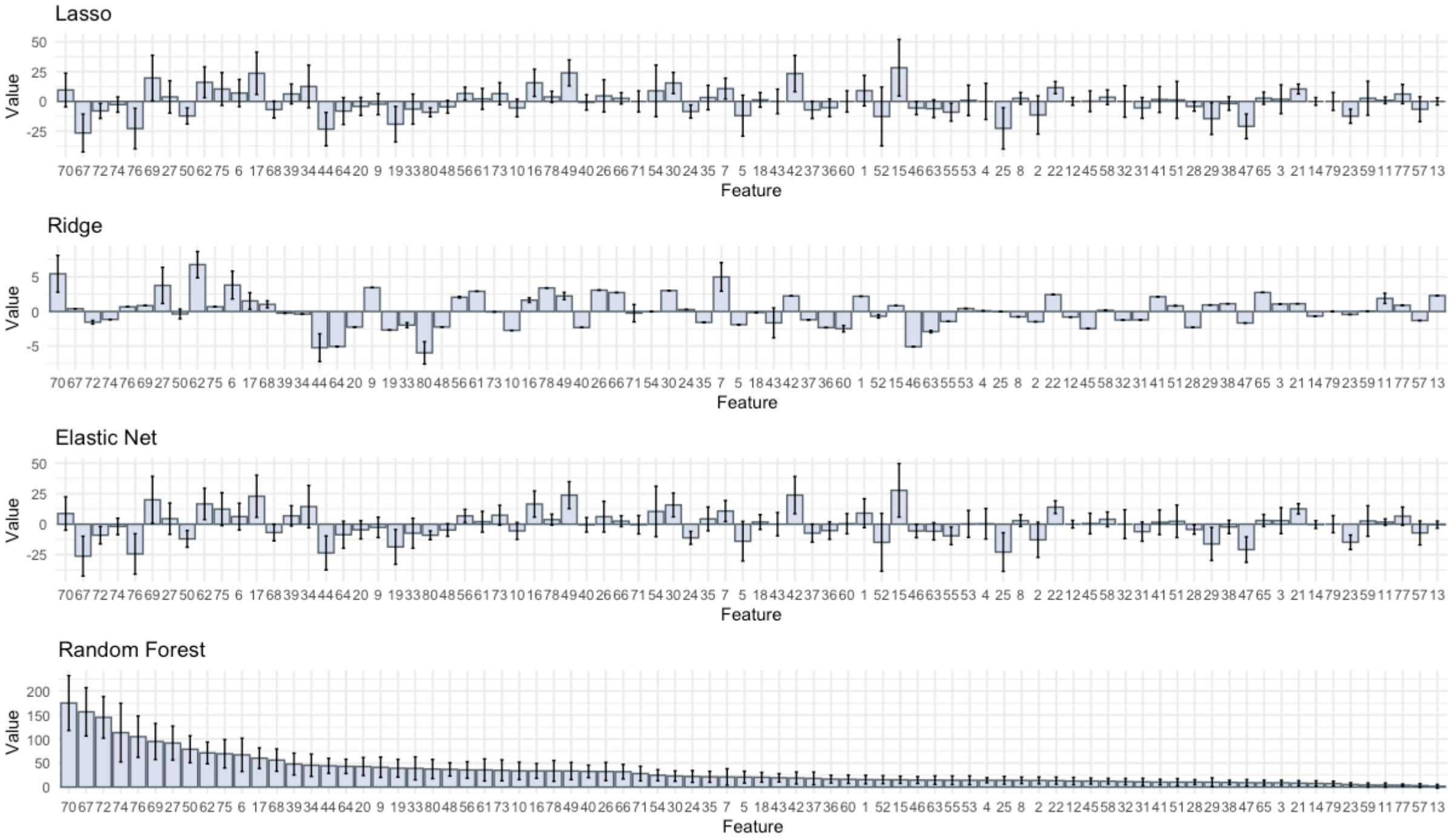


Residuals of Lasso Train



Residuals of Lasso Test

Estimated Coefficients & Importance Of Features



Summary

		Lasso		Ridge		Elastic Net		Random Forest		Multiple Regression	Random Forest	XGBoost	
		Train	Test	Train	Test	Train	Test	Train	Test	Test	Test	Test	
R^2	Min	0.7339	0.6143	0.6869	0.5917	0.7323	0.6233	0.7323	0.7795	0.7274	0.9194		
	1st Qu	0.7469	0.6822	0.7003	0.6613	0.7468	0.6826	0.7468	0.8313	0.7305	0.9225		
	Median	0.7529	0.7149	0.7070	0.6913	0.7517	0.7148	0.7517	0.8456	0.7343	0.9242		
	Mean	0.7534	**0.7115**	0.7071	**0.6905**	0.7530	**0.7123**	0.7530	**0.8472**	0.7400	0.9242	0.9200	
	3rd Qu	0.7597	0.7409	0.7139	0.7180	0.7597	0.7404	0.7597	0.8688	0.7389	0.9260		
	Max	0.7780	0.7966	0.7299	0.7789	0.7752	0.7969	0.7752	0.9034	0.7488	0.9290		
RMSE		18.92		19.59		18.89		13.77		17.63	8.99	9.5	
Top 10 Important Features	1	15		62		15		70				67	range_ThermalConductivity
	2	67		80		67		67				70	wtd_std_ThermalConductivity
	3	49		70		76		72				27	range_atomic_radius
	4	17		44		42		74				64	wtd_gmean_ThermalConductivity
	5	42		46		49		76				69	std_ThermalConductivity
	6	44		64		44		69				76	wtd_entropy_Valence
	7	76		7		17		27				50	wtd_std_ElectronAffinity
	8	25		6		25		50				6	wtd_entropy_atomic_mass
	9	47		27		47		62				72	wtd_mean_Valence
	10	69		9		69		75				44	wtd_gmean_ElectronAffinity
System Time (Fit on whole data)		0.038		0.014		0.016		0.076		Source: Hamidieh, K https://github.com/khamidieh/predict_tc/blob/master/main_script_production_9.R			