Critical Temperature in Superconductors

Richard Kim*
Southern Methodist University

Hien Lam†
Southern Methodist University

Joaquin Dominguez‡
Southern Methodist University

September 7, 2022

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

I. Introduction

Superconductivity was discovered by Dutch physicist Heike Kamerlingh Onnes and his team in 1911.[1] Despite more than one hundred years passing since the discovery, it remains one of the most engaging and enigmatic topics in physics, chemistry, and materials science. As currently understood in simple terms, superconductivity is a collection of physical properties found to exist in certain material classes. Under specific conditions, these materials lose their electrical resistance (infinite conductivity) and, consequently, have the ability to generate very large electric currents and thus, equally large magnetic fields—at least in theory.[2] Modern study of the phenomena in those fields has not proven to be as fruitful, however, given the complex nature of the relationship between the state of superconductivity and the chemical structure of the materials being used.[3]

Engaging with alternative approaches to understanding superconductivity may provide insights not found by and through methods with physical limitations, such as those heretofore employed. Statistical and Machine Learning (ML) methods have advanced enough in recent years to engage with important applications such as this. Specifically, in this case, ML models may predict highly useful information without needing to engage with the chemical structures of materials.

We employed such an approach by exploring the function of critical temperature on more than 160 variables. By using regression methods, such as L1 and L2 Regularization, we were

*Corresponding author
†Corresponding Author
‡Corresponding Author

[1]Superconductivity. CERN. (n.d.).
[2]Combescot, 2022
[3]Chu, Deng, and Lv, 2015

able to create a reliable model that serves to predict critical temperature. As more data-sets related to superconductivity are made publicly available and ML, similar models, in the interest of inter-disciplinary collaboration, may be able to provide much needed clarity to this phenomena.

II. Methods

i. Data Preprocessing

The superconductivity data-set consisted of two files: 'train' and 'material.' The former file contained 82 relevant feature information from 21,263 superconductors including the response feature, 'critical_temp'. The material file contained 88 features which represented the chemical formula for material that were one hot encoded for all 21,263 superconductors. A simplistic example for water (H2O) would have a two under the hydrogen column, one under oxygen, and zero for all the other elements. The two files were concatenated and material subsequently dropped since this feature denoted the full chemical formula and deemed unnecessary after hot encoding. Next, we identified the nine features that had a single value and consequently dropped them due to their lack of usefulness in predictive performance. Lastly, we confirmed that the data was free of missing values and duplicated records. The final unprocessed data frame resulted in 21263 rows and 159 columns. We will note the data-types were all numerical and proceeded with exploratory data analysis.

ii. Exploratory Data Analysis

ii.1 Correlation

Correlations allow us to form a picture of the patterns and dynamics between the variables. We first looked at correlation of features in the following forms:

- Correlation to target variable ('critical_temp')
- Mutual correlation of non-target variables
- Lack of correlation to target variable ('critical_temp')

Identifying those variables that had a high correlation with the target variable provides a safeguard during feature selection. Next, we ran a "for" loop to determine those pairs that show a mutual correlation higher than 0.9 (0 to 1 scale), which gives us an indication which variables affect variance. In this case, 74 different pairs showed a correlation of 0.9 or higher. From those pairs, we concluded that 39 columns show high mutual correlation, which warrant exclusion prior to modeling.

Lastly, we addressed variables with very low correlation with the target variable ($< .01$), which resulted in 69 additional variables being excluded. After these exclusions, our data-set contained 51 variables.

ii.2 Outliers

Upon initial visual inspection of histograms, a few variables displayed skewness with the possibility of outliers.In order to identify and address outliers, and thus reduce MSE, we ran an anomaly detection algorithm ('*IsolationForest*') on the data-set in scaled form, with optimal hyper-parameter tuning. Fortunately, the algorithm only identified two rows as outliers, and said rows were removed.

ii.3 Assumptions

Linearity
In all variables, there exists a linear relationship between the independent variable, x, and dependent variable, y.

Independence
We may assume independence for all instances.

Multicollinearity
Multicollinearity was addressed in the correlation stage of the EDA; those variables with high multicollinearity were removed.

Homoscedasticity
Residuals were shown to have constant

variance.

Normality
Residuals were shown to be normally distributed.

iii. Feature Scaling

Given that the features followed a normal distribution, we felt confident Scikit-Learn's standard scaler class was the most appropriate feature scaling for the dataset. This estimator bounded each feature to maintain a mean of zero and standard deviation of one. We conducted linear regression with scaled and unscaled data and compared their performance by way of five-fold internal cross validation. Looking at R2 and root mean squared error (RMSE), both models produced equivalent metrics at 0.45 and 23.8, respectively. We split the data into X and y (training features and target feature, respectively) and proceeded with standard scaler going forward.

iv. Modeling

As stated above, we would like to predict the critical temperature of a superconductor using linear regression with regularization then investigate the important features. Specifically, lasso and ridge regularization were explored. Briefly explain what lasso and ridge is here. We utilized Scikit-Learn's <i>pipeline</i> class to set up the gridsearch workflow, lasso and ridge classes to train the models, gridsearchcv class to execute the search and refit the best performing model and cross-validate class to ascertain their mean RMSE from five-fold, shuffled internal cross validation. We tuned lasso's alpha to be between 0.1-10 with 200 samples and ridge's alpha to be 0.1-100 with 300 samples. The feature importance was extracted from the best model of each algorithm and its values transformed to the absolute root because we care about the magnitude of the coefficients, not the sign of the coefficient.

Table 1: *Optimal Alpha*

	Optimal Alpha	
	Lasso	Ridge
Alpha	2.03	100
R2	0.66	0.53
RMSE	20.09	22.43

III. Results

i. Regularization

fdfdfdfd

ii. Feature Importance

kjdkfsjkdjkf

IV. Conclusion

i. Code

Please refer to the attached Python notebook.

References

[Chu, Deng, and Lv, 2015] Chu, C. W., Deng, L. Z., and Lv, B. (2015). Hole-doped cuprate high temperature superconductors. *Physica* C: Superconductivity and its Applications, 514, 290-313.

[Superconductivity. CERN, 2009] Superconductivity. CERN. (n.d.). Retrieved September 7, 2022, from https://home.cern/science/engineering/superconductivity

[Combescot, 2022] Combescot, R. (2022). Superconductivity: An Introduction. Singapore: Cambridge University Press.

[Figueredo and Wolf, 2009] Figueredo, A. J. and Wolf, P. S. A. (2009). Assortative pairing and life history strategy - a cross-cultural study. *Human Nature*, 20:317–330.