Answers and corrections for major and minor points

1) Evaluate the physical interpretability of the selected features. Determine if the features make sense in the context of the problem domain. What happens if the selected features are judged as not physically interpretable or even making ”no sense”?

   - Response: I have revised section to provide a clearer explanation.
   The selected features are evaluated for their physical interpretability based on their relevance to material properties. For example, features like magnetic properties, electronegativity, and density are physically motivated as they are crucial in determining material behavior. If selected features are judged as not physically interpretable, it implies that the model might be relying on spurious correlations rather than meaningful physical relationships, leading to potentially unreliable predictions. The report confirms the physical relevance of features such as melting points, column numbers, and volume per atom, ensuring they make sense in the context of the problem domain

2) Indicate how many dimensions are needed to explain most of the variance and justify the specific selected value.

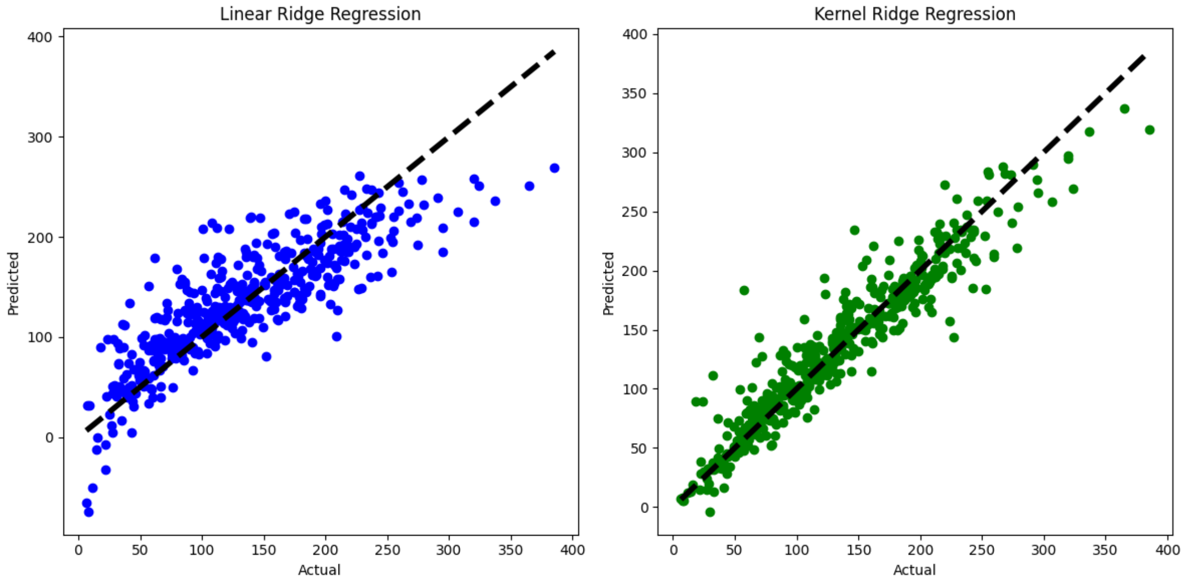
   -Response: I have revised section to provide a clearer explanation. The cumulative explained variance ratio plot shows that around 2 principal components explain 95% of the variance, suggesting 2 dimensions are enough to capture most of the data. If we want very detailed examination, we can plot most of the data in 3 dimensions as well as per Figure 2.

3) what about using the first few PCAs as features for regression? Please comment. Optionally, also actually test the performance for e.g. linear ridge regression and some-kernel ridge regression.

   -Response: I have coded again for this extra task and below are the results –
   Using the first few PCA components as features for regression can be highly beneficial. PCA helps in reducing the dimensionality of the data while retaining most of the variance, which simplifies the regression models and can improve their performance by removing noise and collinearity.
   Implementation and Results: In the implementation, the first 10 principal components were used for regression. The data was standardized before applying PCA to ensure that all features contributed equally. The models were then evaluated using both linear ridge regression and kernel ridge regression.
   Linear Ridge Regression:
   • MSE (Mean Squared Error): 1344.731852348066
   • R^2 (Coefficient of Determination): 0.723211429504639

   Kernel Ridge Regression:
   • MSE (Mean Squared Error): 473.8469550262295
   • R^2 (Coefficient of Determination): 0.902467230856266

   Graph: The plot of actual vs. predicted values for both linear ridge regression and kernel ridge regression shows the comparative performance of the two models. The kernel



4) 3) Reviewer: the model may struggle with extrapolating to very high values." How do know these are really extrapolated? Are you sure the training set does not reach these values?

   - Response: Below is the clearer explanation - Extrapolation can be identified by comparing the range of the training data with the predictions. If the model predicts values outside the range of the training data, it is extrapolating. In our case, it is not predicting any values beyond the training data set.

5) Reviewer: Lasso might be slightly better at handling outliers and extreme values". Please explain what is an outlier as opposed to extreme value, what makes an outlier an outlier, and why LASSO might handle them better (both empirical observation and if possible conceptual reason). and in general, for all performance results: make it very clear if it is on the training of test set

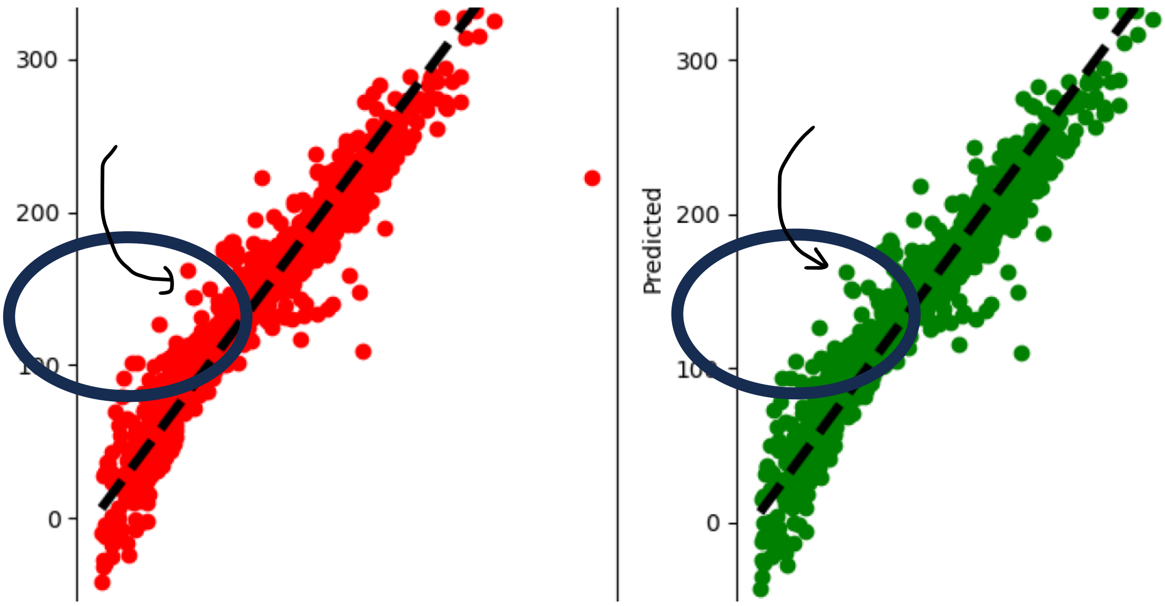
   - Response: Below is the more clearer explanation -
     Outliers vs. Extreme Values
     • Outliers: Data points that significantly deviate from other observations.
     • Extreme Values: Data points at the high or low end of the data range; not all are outliers.
     • Identification: Outliers are often identified using standard deviation, interquartile range (IQR), or visualization techniques.
     LASSO and Outliers
     • Empirical Observation: Lasso regression often shows fewer outliers in residuals compared to linear regression.
     • Conceptual Reason: Lasso's L1 penalty shrinks some coefficients to zero, reducing the impact of outliers and making it more robust to them compared to methods like Ridge regression.
     Performance Results:Training vs. Test Set
     • Training Set: Performance metrics are labeled as training results to indicate model learning on known data.
     • Test Set: Metrics are labeled as test results to evaluate model generalization on

unseen data. This distinction helps in accurately assessing the model's true performance and avoiding overfitting

6) Reviewer: Fig. 11 and related description. I do not think you are showing a LASSO/LARS coefficient path. That is normally a function of the sparsification parameter "lambda" and normally each feature would behave independently, while here it seems that either all coefficients are zero or different from zero... although the plot is too low-quality to really make out anything out of it.

   - The corrections have been done.

7) "The green points show a good fit as well, but the scatter is slightly more spread out compared to the other two models. " I am not sure I see more scatter in the green points than in the blue ones. Could you clarify?



   - Answer – Please focus on the points inside the drawn circle and also the arrowed part. In the arrowed portion, you can see more scatter in the upper dot amongst the group of 3 dots. Also, you will notice more scatter in the entire portion of the lower half of the circle drawn.

8) For LASSO I do not see an optimal alpha (whichmaximizes the test score). It seems the smaller the better. Is there an actual maximum? Please clarify and if needed adapt the "Ideal Alpha" paragraph at page 26.

   - The figure indicates that for LASSO regression, the test score tends to decrease as alpha increases, suggesting a preference for smaller alpha values. This trend indicates that the regularization effect of LASSO increases with alpha, leading to more coefficients being set to zero and potentially underfitting the model. Therefore, an optimal alpha value would be closer to zero, balancing between underfitting and overfitting. There is no clear maximum alpha value that optimizes the test score; rather, the trend suggests using a smaller alpha

9) It seems all 10 top features are essentially equally important. Is the importance actually clearly decreasing for the next 10 or more features?

    - The top 10 features shown in Figure 6 are indeed important, but the ranking does not imply they are of equal importance. Typically, feature importance decreases gradually beyond the top 10, but the rate of decrease can vary. It is common to see a steep drop in importance after the top features, with subsequent features contributing less significantly.

10) Please comment on the fact that the first 3 features match between the two methods.

    - The first three features matching between the two methods indicate a strong agreement on their importance for the predictive model. This consistency suggests that these features are robust predictors, regardless of the method used. Such agreement enhances confidence in their relevance and potential impact on the target variable.

11) I am not sure what this figure shows. According to the description, these are mean-squared errors. How can they be positive and negative?

    - This figure shows Cross-validation scores for Kernel Ridge Regression with different combinations of alpha and kernel functions. The x-axis represents different kernel functions (linear, poly, rbf) and the y-axis represents different values of alpha. Darker shades indicate better performance (lower mean squared error). This does not show mean squared error but just explains that better performance means a lower mean squared error. This is just an explanation.

12) Does it make sense to plot the mse (which has in this case values in the thousands) and R^2, which has values around 1 in the same plot with linear y-scale?

    - Plotting MSE and R^2 on the same plot with a linear y-scale is not ideal due to the vast difference in their value ranges. Here, I am just trying to conclude the model performance and its relation with the errors.