Question 1

What is the optimal value of alpha for ridge and lasso regression?
What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Alpha (α) is a penalty term that indicates the amount of shrinkage applied in the equation. If the alpha is set to zero, it is equivalent to a linear regression model and the larger value penalizes the optimization function. The optimal values of alpha for ridge and lasso regression depend on the data and the purpose of the study. One way to get the best alpha is to use cross-validation. That is the best way of checking how a model performs on unseen data. The experiment with alpha values and the chosen data that minimizes cross-validation error. If twice the alpha value is selected for the ridge and lasso, the changes to the model will be as follows.

Ridge regression:
The coefficients will be smaller in magnitude, because the penalty term will be larger. This will reduce the variance in the model, but will also increase the bias. The most important predictive variables will be the ones with the greatest degree of accuracy, as they will have the greatest influence on the outcome variables.

Lasso regression:
The coefficients will be zero, because the penalty term will be large. This will result in larger feature selection, as some predictor variables will be removed from the model. The most important predictor variables will have non-zero coefficients, as they will have the greatest impact on the outcome variables.

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

The choice between Ridge and Lasso regression depends on the nature of the data and the objective of the analysis. The optimal lambda values are different for individual error metrics. The best metric is considered to align with your goals for the model. Here are some things to consider,

Ridge regression:
This is best suited when there are many predictor variables that are correlated with each other, because it will reduce multicollinearity problems. This includes all predictor garbage in the model, that can lead to overfitting if there are many features. This has a closed form solution, that can be solved analytically.

Lasso regression:
This is more suitable when there are many predictor variables that are irrelevant or unimportant, as it will feature selection by some establishment over coefficients to zero. This automatically selects the most important features, that can lead to better generalization and interpretation. This does not have a closed form solution, that finding the optimal coefficients requires an iterative algorithm.

This can affect the calculative efficiency and stability of the models. So the best approach is to compare the performance of the two models on the data by analytical metrics like mean square error, R-square, or cross-validation score.
Coefficients of both the models can be plotted to check to find the uniqueness.
Based on the results, the model can be selected that best fits your data and meets your objectives.

Question 3

After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data.

You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?
The best way to answer is to fit a new lasso model to the data that does not have most important predictor variables. The five most significant predictor variables are those with zero coefficients in the original lasso model. The coef_can be used as property of the Lasso object to obtain coefficients. After inserting a new lasso image, the coefficients can be verified to see which ones are not zero. These will be the five most important forecast variables in the new model. Alternatively, you can use the select From Model function from the sklearn.feature_selection module to select the same features with zero coefficients.

Question 4

How can you make sure that a model is robust and generalizable?
What are the implications of the same for the accuracy of the model and why?

A model is robust when it can perform well on unobserved data that may have different characteristics or distributions from the training data. Some common strategies to ensure that a model is robust and generalizable are,

Cross-validation:
This method of dividing the data into several groups. Some of them are used for training and others for testing. The model can be evaluated on different subsets of data and the average performance can be calculated.

Constant:
The method to add a penalty term to the loss function of the model to prevent overfitting. This can reduce the complexity of the model and increase its fit to new data. There are different regularities, such as L1 (lasso), L2 (ridge), or drop out.

Data enhancement:
This method of increasing the size and variety of data by applying some transformation, such as cropping, flipping, rotating, or adding noise This can make the model more robust to transformation of the input data to improve its generalization.
The definition of robustness and generalizability for model accuracy depends on the trade-off between bias and variance. Bias is error due to assumptions or simplifications in the model, while variance is error due to data sensitivity of the model. A more robust model may have low bias but high variance, that is, it may fit the training data well but perform poorly on new data. A very simple model

may have high bias but low variance, that can operate consistently but does not capture the underlying structure of the data The best model is the one that balances bias and variance and has the highest accuracy on the training and testing data.