Answer 1: The optimal value of alpha for both, lasso and ridge regression are generally decided by using techniques like cross-validation. In the assignment, a grid search cv is used to find the best alpha. If we were to increase the value of alpha, for ridge, the variables would tend to move towards 0 and for lasso, as the alpha value increases, the variables could get to exact 0. So, if we double the alpha value, in ridge, the model will regularize more and the variables would be pushed towards 0 and for lasso, the sparsity will increase which will set more variables to 0. This approach could result in reducing overfitting but could also increase the chances of underfitting. If the values are doubled, in ridge, the variables which are already less importance could well go towards 0, further decreasing the importance while in lasso, variables which don't have 0 coefficient would get highest weightage.

Answer 2: Ridge and lasso, both, are regularization techniques techniques which would reduce overfitting however they would do it in their own different ways. Richwood typically add a penalty to the square of the coefficients while lasso would had a penalty to the absolute value of coefficients. We could typically say that a lasso model would tend to produce sparse models by taking the coefficients to exactly zero thus it would perform an effective feature selection. If we were to choose between the rich and lasso we would consider the following aspects:

i. Feature importance and sparsity: If our use case requires most of the features that we need in the model would be relevant we would typically go with ridge as it shrinks coefficient towards zero but rarely reduce in them to exactly zero. If we believe that our model is sparse and not all features would be impacting the final outcome we might want to go ahead with lasso as it would typically use lesser variables which would help in interpretation of the model.

ii. Multi - collinearity: A Ridge model could be effective when we deal with multicollinearity (higher correlation between independent variables), as it would distribute the weights amongst all variables in the case of lasso where only selective variables are used in the model it would typically choose one variable when and if there is collinearity observed.

iii. Interpretability: If our use case is to have fewer variables and better understanding of the model in other words better interpretable model we would go ahead and use lasso as its ability to make coefficients exactly zero would help in the interpretability of the model

Answer 3: In such a scenario, we would not consider the top 5 variables/features used in the lasso model. We would go back and re-run the model without those 5 features. The best features from the 2nd run would be the next best 5 features from the model by ranking the features with absolute coefficient.

Answer 4: To ensure robustness and generalizability, we would use Cross validation on train and test splits to understand the performance on multiple subsets of data. We would use regularization techniques like Lasso and ridge to prevent overfitting. Intensive feature engineering to keep only relevant features in the model, excluding outliers. Interpretability might hamper accuracy to a point, so a balanced tradeoff should be taken care of