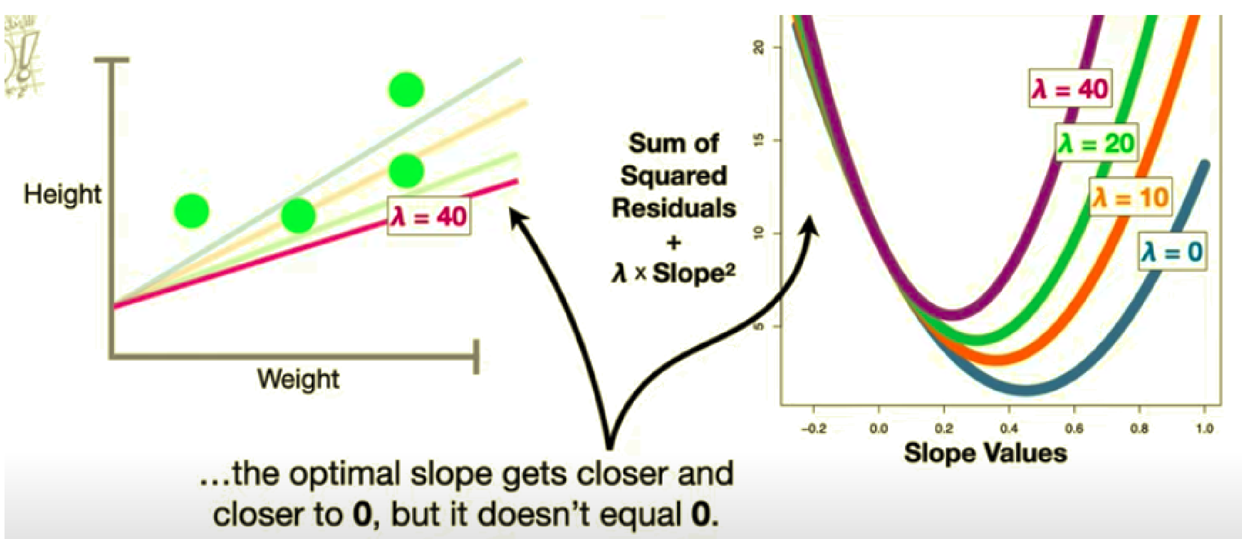
Ridge vs Lasso Regression Explained

NORMAL LINEAR REGRESSION

- Normal regression gives you unbiased regression coefficients (maximum likelihood estimates “as observed in the data set”
- We use the regularization techniques to avoid overfitting on unseen data by numbing, muting, or adjusting the coefficients of our original linear regression .
- To find the optimal values for ridge and lasso regression we can use cross validation and find the best parameters to shrink our given coefficients for our independent features.

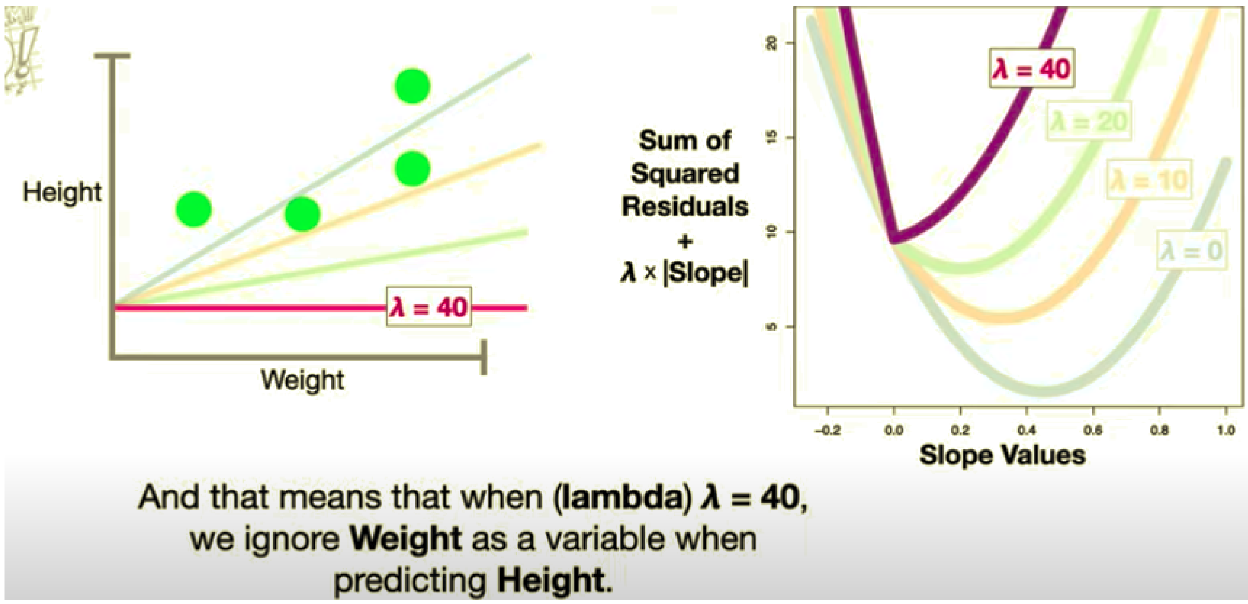
RIDGE (L2) alpha is the same as lambda   - lambda x slope squared

- Ridge regression is used when the data set contains a higher number of predictor variables than the number of observations, or when multicollinearity is present.
- Ridge minimizes the sensitivity towards independent variables through the parameter lambda.
- When lambda equals zero the penalty equals zero regardless of the slope, and leaves us with our original sum of squared residuals.
- As lambda increases our slope decreases effectively decreasing the sensitivity of our independent variables in our linear regression algorithm
- Optimizing lambda lies within the equation as we want to find the bottom of the parabola or the lowest sum of square residuals plus penalty.
- A general note to state is the optimal slope will never equal zero meaning the features will never be completely eliminated.



LASSO (L1) - lambda x |slope|

- Lasso removes variables by setting their coefficients to zero.
- When lambda is zero the penalty is zero leaving us our original sum of squared individuals.
- When lambda grows to a given amount we can completely ignore a given independent variables all together effectively removing it.



Test Answer

When finding the optimal training line for a Linear Regression model, the presence of variance in the test set is inevitable. In the absolute simplest way possible, variance refers to how spread out the data may be given the mean of the set. Because of this issue, the regularization parameters of Lasso and Ridge are tools which apply penalties to our residual points. The mathematical formulas are very similar to each other, but differ on one simple fundamental. Ridge regression penalizes the sum of squared coefficients, and Lasso penalizes the sum of their absolute values.

Inherently, because of this fundamental difference Lasso narrows down the features with its formula, and for that reason it performs better when you have only a handful of features which have greater impact on the model. Lasso does a good job of eliminating or minimizing certain features which already had a significantly small variance to begin with. Hence the name Least Absolute Shrinkage and Selection Operator (LASSO), because the formula provides a narrow focus on the more impactful varying data, and penalizes them appropriately, it is a good method to choose when there are less significant features.

Ridge regression keeps all features intact, and applies the penalization parameter in a manner where all features are still represented. Therefore Ridge is the best choice when there is a dataset with many parameters that are very similar to each other. They have a high multicollinearity value, and their impact in predicting one another is fairly consistent.

LESS SUITABLE ALGORITHM (K MEANS CLUSTERING)