- What makes a model linear? Linear in what?
 When the independent and dependent terms have a linear relationship(makes a line rather than a curve) in the expression.
- 2. How do you interpret the coefficient for a dummy/one-hot-encoded variable? (This is a trick question, and the trick involves how you handle the intercept of the model.)

The coefficient for a dummy variable represents the average difference in the dependent variable between the group represented by the dummy variable and the reference group.

3. Can linear regression be used for classification? Explain why, or why not.

Linear regression predicts continuous values, but classification requires discrete values.

4. What are signs that your linear model is over-fitting?

High variance is one sign that the model is over-fitting meaning the model is sensitive to small changes in the training data. Also, if there is a big difference in performance (accuracy) of the training and testing set. For example, if the training set is highly accurate, but then the model does not perform well on new data, this indicates that the model is likely over-fitting.

5. Clearly explain multi-colinearity using the two-stage least squares technique.

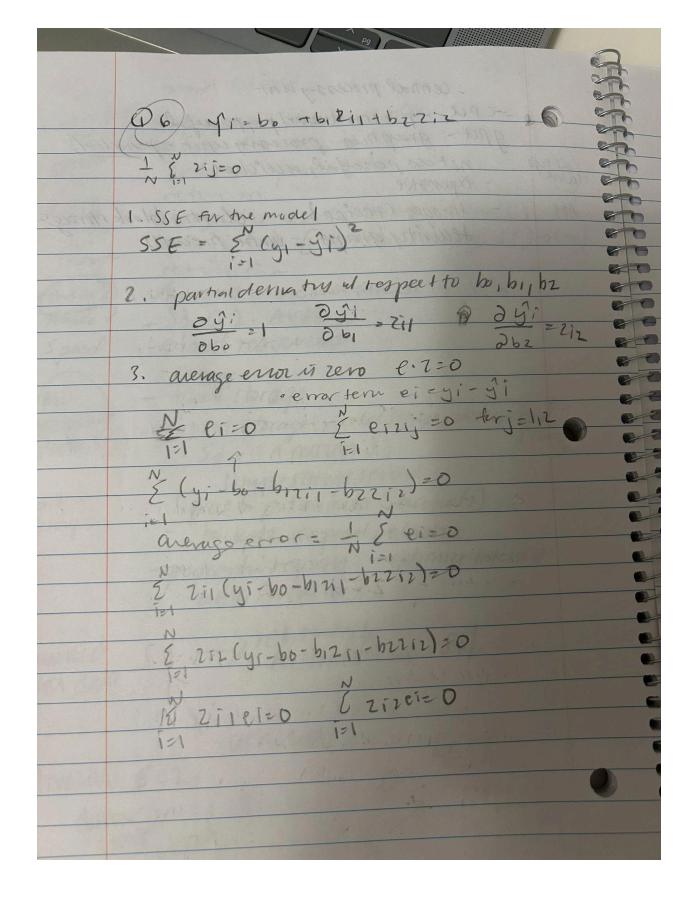
Multicollinearity is when the independent variables are highly correlated. The two stage least squares technique can minimize the effects of this problem. In the first stage, you regress the endogenous variable on the exogenous variable. The goal here is to predict an endogenous variable based on the exogenous one to predict a "clean" value. In the second stage, the dependent variable is regressed on the predicted values (from stage one). The goal here is to estimate the relationship between the dependent and predicted variable to lessen the impact of multicollinearity.

6. How can you incorporate nonlinear relationships between your target/response/dependent/outcome variable y and your features/control/response/independent variables x into your analysis?

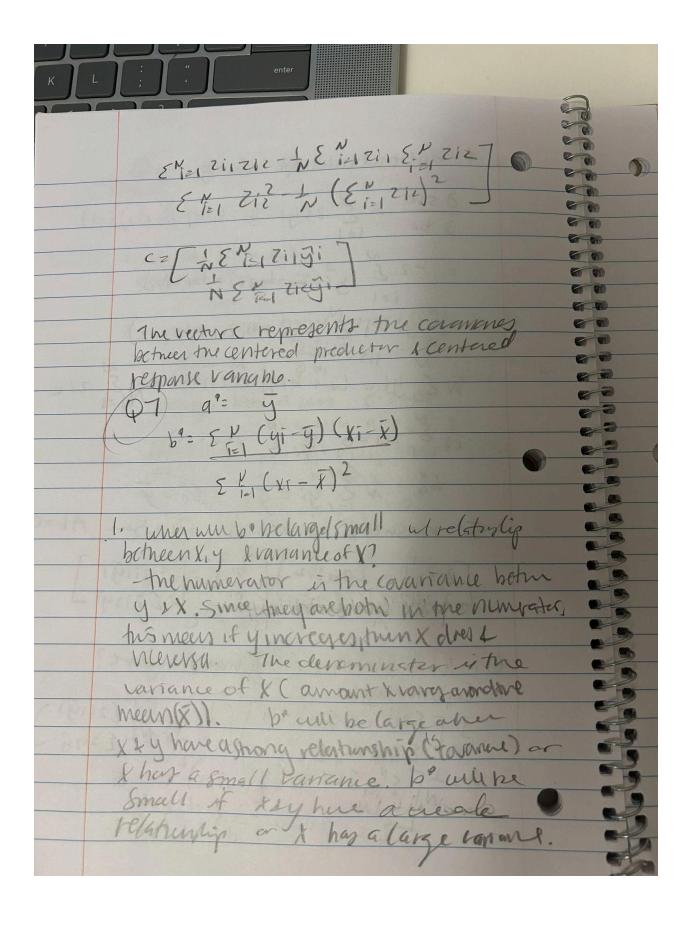
One way to incorporate nonlinear relationships between x and y variables is to use tree models like random forests, which can capture complex nonlinear relationships.

7. What is the interpretation of the intercept? A slope coefficient for a variable? The coefficient for a dummy/one-hot-encoded variable?

The intercept is the predicted value of y when the x is zero. The slope is the value of a one unit increase in x. The coefficient of a dummy variable is the difference of the mean of the dependent variable between the category and reference category.



4. optimel intercept 60=5 ∂ 55€ = & 2(y1-b6-b1211-b1212)(-1) 5. results as matrix equation Ab= 6. What is matrix A? what is vectors? Explain inturtion.



7. affect interrept of regression? What nappen to be coefficient? How densit affect ability to predict? - The exercis independent, 50 the interrupt and not be affectelby measurement ener - when there is measuret error, 50 uil ferel to be biased toward zero, · The ability to predut will usen Therewin be greater uncertainty 3. noise ni is independent - Barrially, the note does not introduce bias (bic stress mean tindependent), but it does reduce precision of our estimate of the relationship of X & Y. The added noise of & inflator the variance. This leads to attenuation of 6" . Attenuation is important in the cost-benefit analysis of data quality. unen musnement error is large, clearer data/ acuate measurents are largely beneficil for predictions. Alternatively, if the medsnement error in small, more effort to den data to archier, not north the cost.