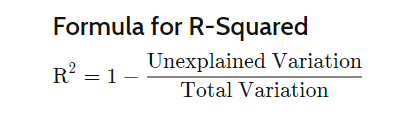
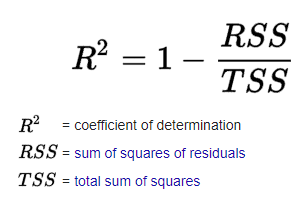
Short question interview

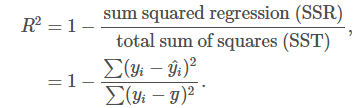
* Speak enough
* Don’t go thru other points if not prompted

What is R2? What are some other metrics that could be better than R2? And why?

R-Squared is a statistical measure of fit that indicates how much variation of a dependent variable is explained by the independent variable(s) in a regression model.







Residuals measures the difference from the actual and the predicted value (dependent variable)

Total sum of squares takes the sum of difference between predicted and the mean y

Generally the higher the R2, the lesser the variation of y values accounted by each x values.

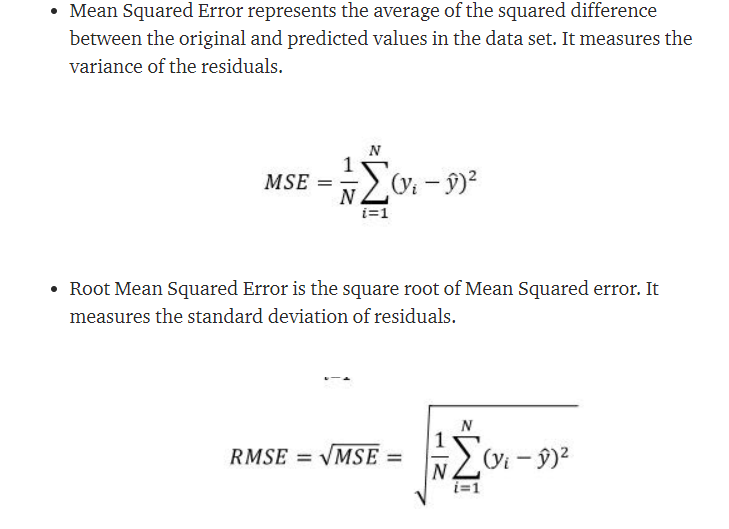
R2 = 0, higher variance, the model cannot account for the y values.

What metrics can be better than R2?

For comparing the accuracy among different linear regression models, RMSE is a better choice than R squared.

**RMSE** tells us how a regression model can predict the value of response variable in absolute terms

* We can use MSE to tell whether or not the model is over or underfitting based on its difference between the predicted(test) and the cross val score(train) of the linear regression model



While

**R2** tells how well the predictor variables can explain the variation in the response variable.

What is the curse of dimensionality?

It means the error increase with the increase in number of features. This overfitting can be reduced by two ways,

1. Dropping features – loses any benefit from those we dropped
2. Use PCA – get benefits of most of the features

To reduce this, we need to apply dimensionality reduction methods

`Feature selection is to drop variables from our model`

`Feature extraction -- \*\*another form of dimensionality reduction\*\* -- takes existing features and combine them together linearly, making new variables called Z. It helps to reduce the number of features while still keeping the most important pieces of original features`

PCA reduces the number of variables of a data set while preserving as much information as possible.

Calculate the covariance matrix for all variables with one another. Decompose it to eigenvectors. The eigenvalue is a number that tells us how important each direction is. Eigenvalues will be sorted from biggest to smallenet.

Using explained variance

Principal component analysis is a method of dimensionality reduction that **\*\*identifies important relationships\*\*** in our data, **\*\*transforms the existing data\*\*** based on these relationships, and then **\*\*quantifies the importance\*\*** of these relationships so we can keep the most important relationships and drop the others!

PCA distorts interpretability!!!

Is more data always better?

Set the bounds the question

I am going to interpret “better” as “building a better predictive model”

* Bias variance tradeofff
* Prediction risk breaking down into bias, variance and irreducible error
* Context and examples

I am going to interpret “more” as more columns only

Can be about rows.

Generally the more data we have the better, as we can have more features (columns) or better training. In the context of a neural network, feeding in more pictures of dogs vs cats, will help the machine learn more by digesting more variations so as to call a better prediction.

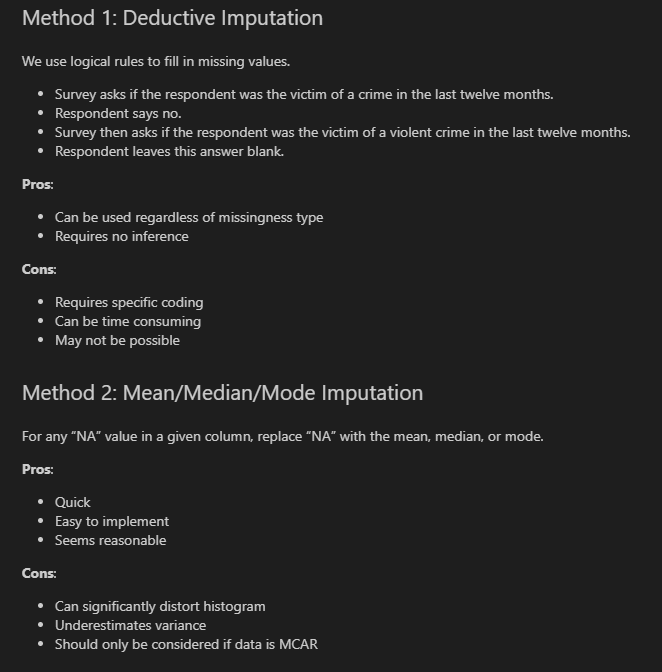
It is not always better, if lets say model is too simple, hence causing high bias. So even throwing more data will not make things better. For example for the case of CNN, if we have a model that has very few layers, the hidden notes will not be able to pick up relationships.

How do you deal with some of your predictors being missing?

Missing completely at random

Missing at random (condition on data we have observed) – can see some sort of systematic

Not missing at Random



Predictors = features

1. Gather resources to obtain the data
2. Work with what we have

Missing predictors = missing values

1. missing completely at random 🡪 remove the observations
2. Not random 🡪 inpute a data in relation to other features
3. Not missing at random 🡪 traffic light 8-10am not here

How to add structure

Question 12) Now you have a feasible amount but youre fairly sure that youi don’t need all of them. How would you perform feature selection on the dataset?