The most common problems that can occur while training a ML model

AI BUSINESS SCHOOL

Underfitting

When data lacks complexity

Overfitting

When data is too complex

When does underfitting happen?

AI BUSINESS SCHOOL

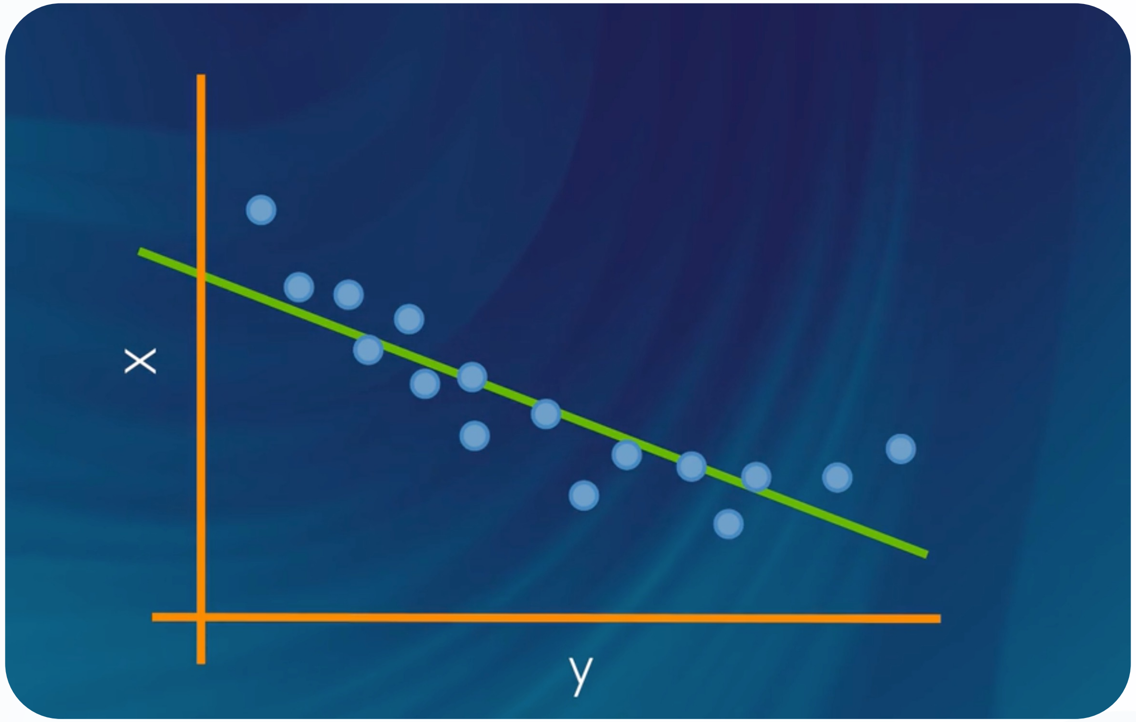
- A model is too simple or lacks complexity
- A model is unable to find the patterns in the training data
- A model generates a high error on the training set & unseen data

The inability of the model to understand complexity of data: bias

Underfitting models can also be referred as **“highly biased”:**

- A very simple straight line that does not fit the data properly
- A large portion of the dataset is ignored

▶ The model performance is poor



When does overfitting happen?

AI | BUSINESS SCHOOL

- The model is trained too much on a specific training dataset
- The training data is very specific and has too many features
- The model is unable to generalize testing data; showing low accuracy

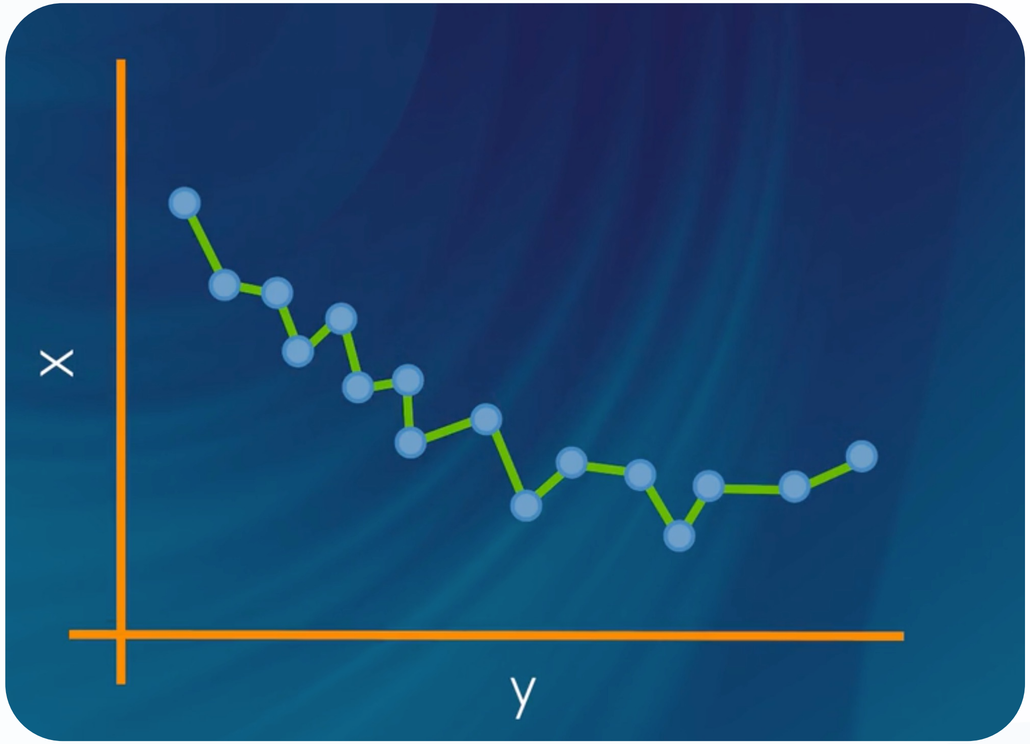
The sensitivity of a model to a specific dataset: variance

AI BUSINESS SCHOOL

Overfitting models can also be referred as **"high variance models":**

- A very complex line is fitting each datapoint but fails to recognize the general pattern

▶ The model is unable to make accurate predictions on new data



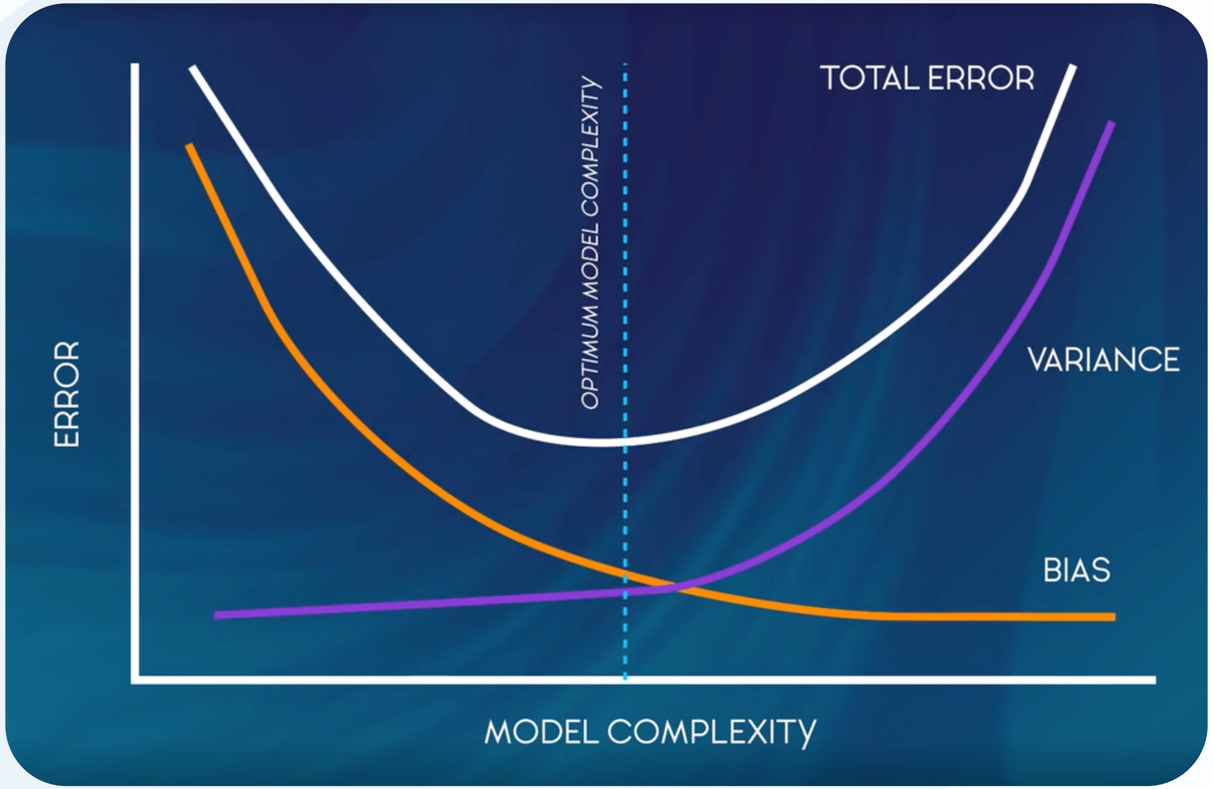
The aims is to achieve a good balance between the bias and the variance

- The performance of the model is affected by both variance and bias which can lead to underfitting and overfitting and eventually cause poor predictions.
- By adjusting variance and bias, we can generalize the model so that it is neither **too complex** nor **too simple**.

The trade-off between bias and variance

AI | BUSINESS SCHOOL

- As variance **increases** bias **decreases**
- As bias **increases** variance **decreases**



How can we solve overfitting and underfitting?

AI BUSINESS SCHOOL

To solve underfitting

making the data **more complex** by increasing the number of observations in the training set & adding new features

To solve overfitting

making the data **less complex** by removing complexities

We can use regularization to reduce complexity

How does regularization work?

Regularization shrinks coefficients **towards zero**, so that the impact of less significant features is **reduced**, and high variance is prevented.

Regularization uses loss functions: L1 and L2

L1

- Used in lasso regression
- Less common
- Not affected by outliers as it is just considering the difference between actual and predicted values

L2

- Used in ridge regression
- More common
- Not useful on dataset with outliers as it is taking the squared difference which will increase the error