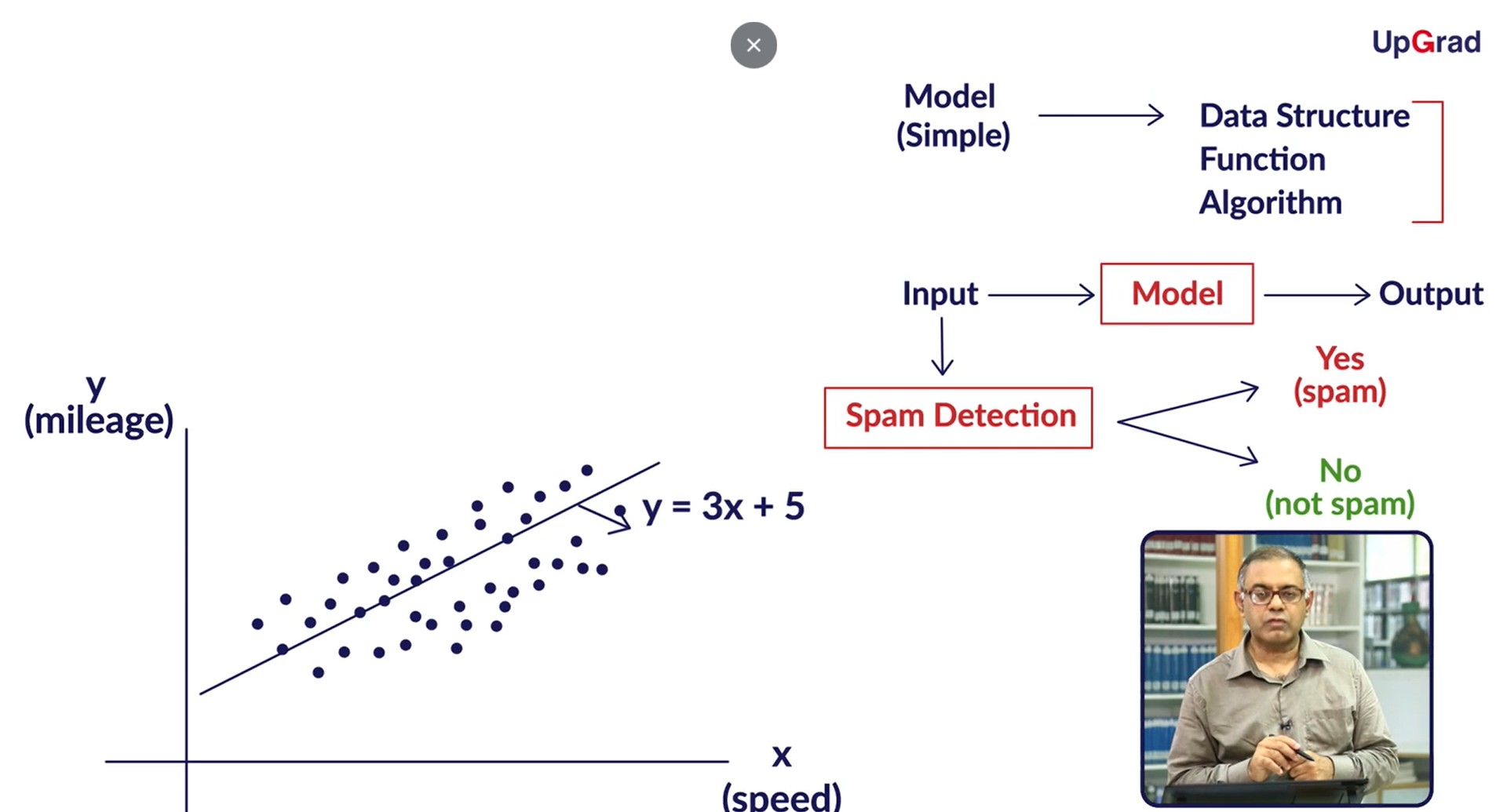
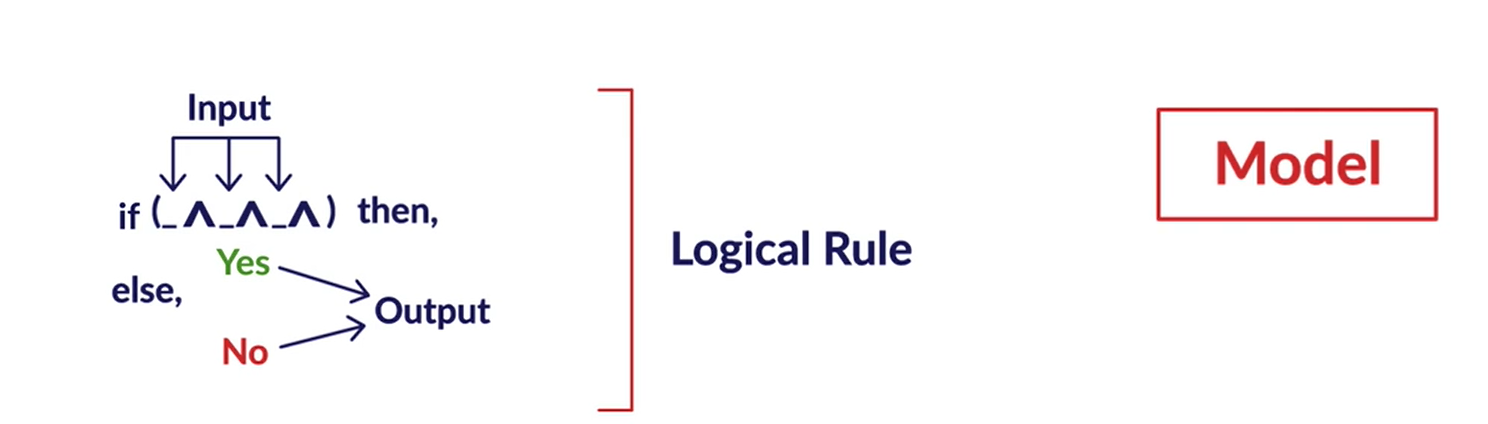
**Principles of Model Selection:**

* **Introduction to Model Selection.**
* Hum log bohot saare models banake unka results compare karte hai aur fir select karte hai optimum model. But can we try all possible models? It is not possible right? Uske liye kuch principles follow karne padte hai.
* Uske alava domain knowledge bohot important hai. Uske basis pe hi model selection depend karta hai, testing parameters and methods, kis tarah ke trade offs acceptable hai.

Occam’s Razor: Making your model as simple as possible but, no simpler. If the model made is more complex than it should, then there is a problem at hand. If the model made Is simpler than it should, then it is too weak to get expected results.

Usually there’s a confusion that a model is definitively always an algorithm. Is it? No. A model is anything that takes input, processes it(trains it, or if else process etc) and gives us satisfactory output. 



Simplicity can unanimously decided without putting any definition to it but let’s say I cases of decision tree it can be said that the model with less depth or less number of nodes is a simpler one. In terms of linear regression: Higher degree of polynomial is more complex. In case of both SLR: using coefficients and number of bits we can decide.

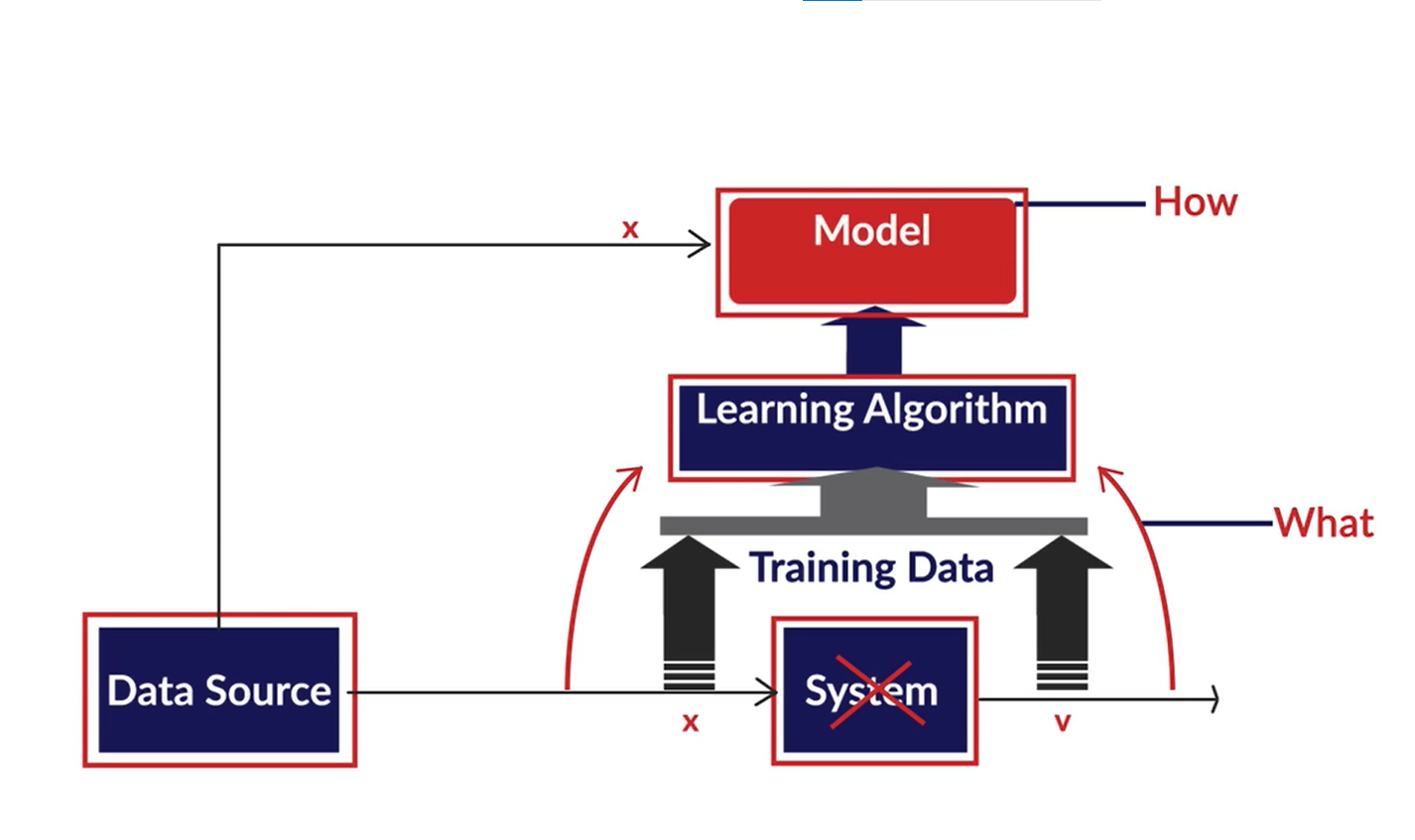


Figure 1/**Basic ML Framework:**Here the system is the Email police viz replaced by a model developed using Learning Algorithm.

* **Simplicity, Complexity and Overfitting:**

There are a few practical issues that we face while making a decision about the type of models too implement before going into solving business problems.

1. Type of data: Numerical, categorical, date-time, string etc.
2. Data Quality: Missing values, Noisy data.
3. Dimensionality of data: High or Low.

One more assumption that’s burst in this video. Comparison of models must never be on training data. It always has to be on unseen data on which the models are not built on.

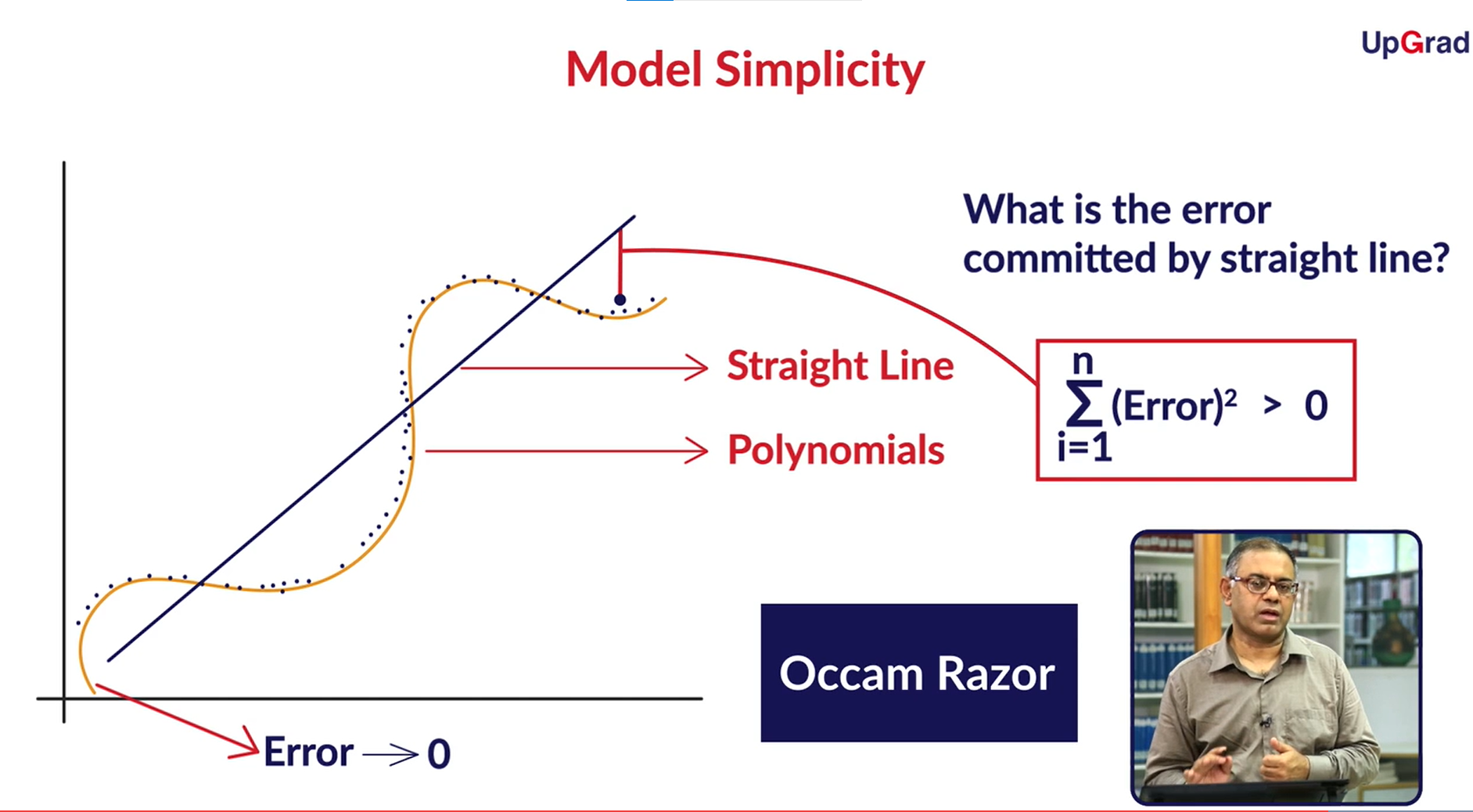
How to choose a model from same sort of problems?

Figure 2/ Regression problem. We can either choose from the class Straight line or from the class polynomial. Using Occam Razor we must go for Linear class unless we have enough evidences to prove that the results by simpler model is not satisfactory.

Complex vs Simple models.

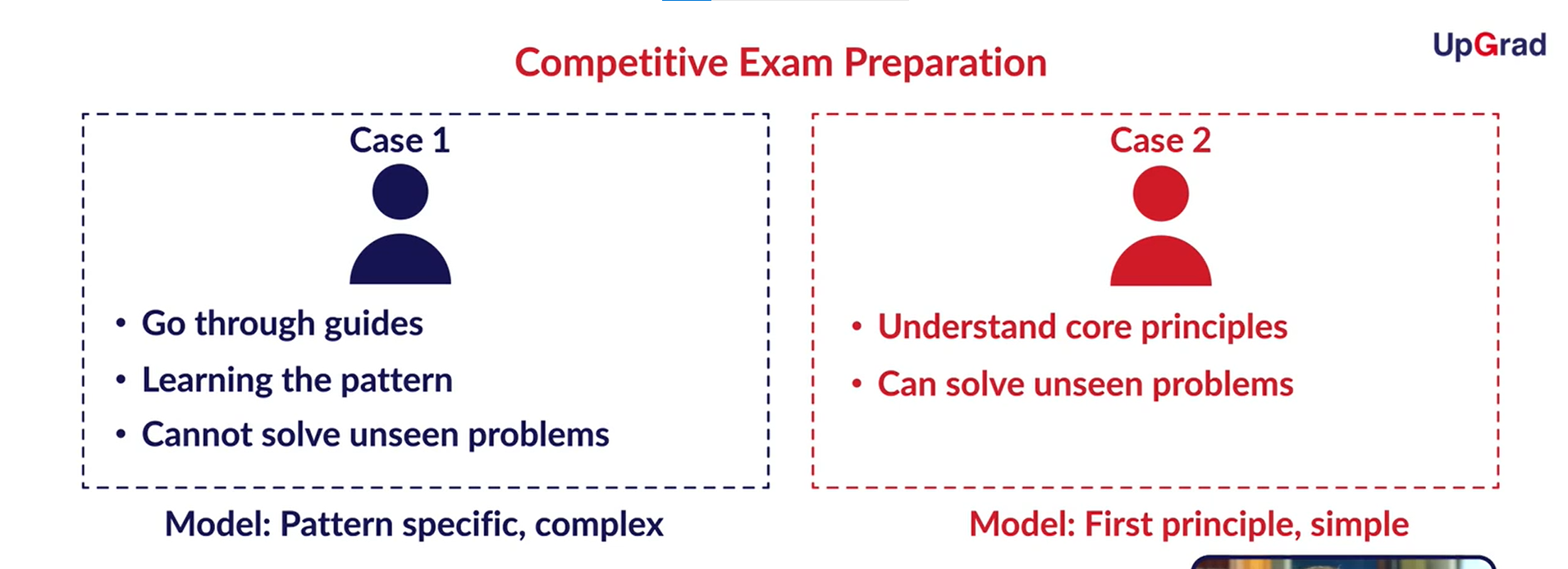


Figure 3/ Figure to understand why Simple models are better than Complex models.

1. Simpler models are usually more generic: Figure 3
2. Simpler models require fewer samples. Complex pattern: Solve many guide books to understand patterns.
3. Simpler models are more robust: Person on the right doesn’t care if the pattern of the exam changes in the last minute. Whereas for the person in left, his/her world will be upside down.

Conclusion:

1. A simpler model is usually more generic than a complex model. This becomes important because generic models are bound to perform better on unseen data sets.
2. A simpler model requires fewer training datapoints. This becomes extremely important because in many cases, one has to work with limited datapoints.
3. A simple model is more robust and does not change significantly if the training datapoints undergo small changes.
4. A simple model may make more number of errors in the training phase, but it is bound to outperform complex models when it processes new data. This happens because of overfitting.

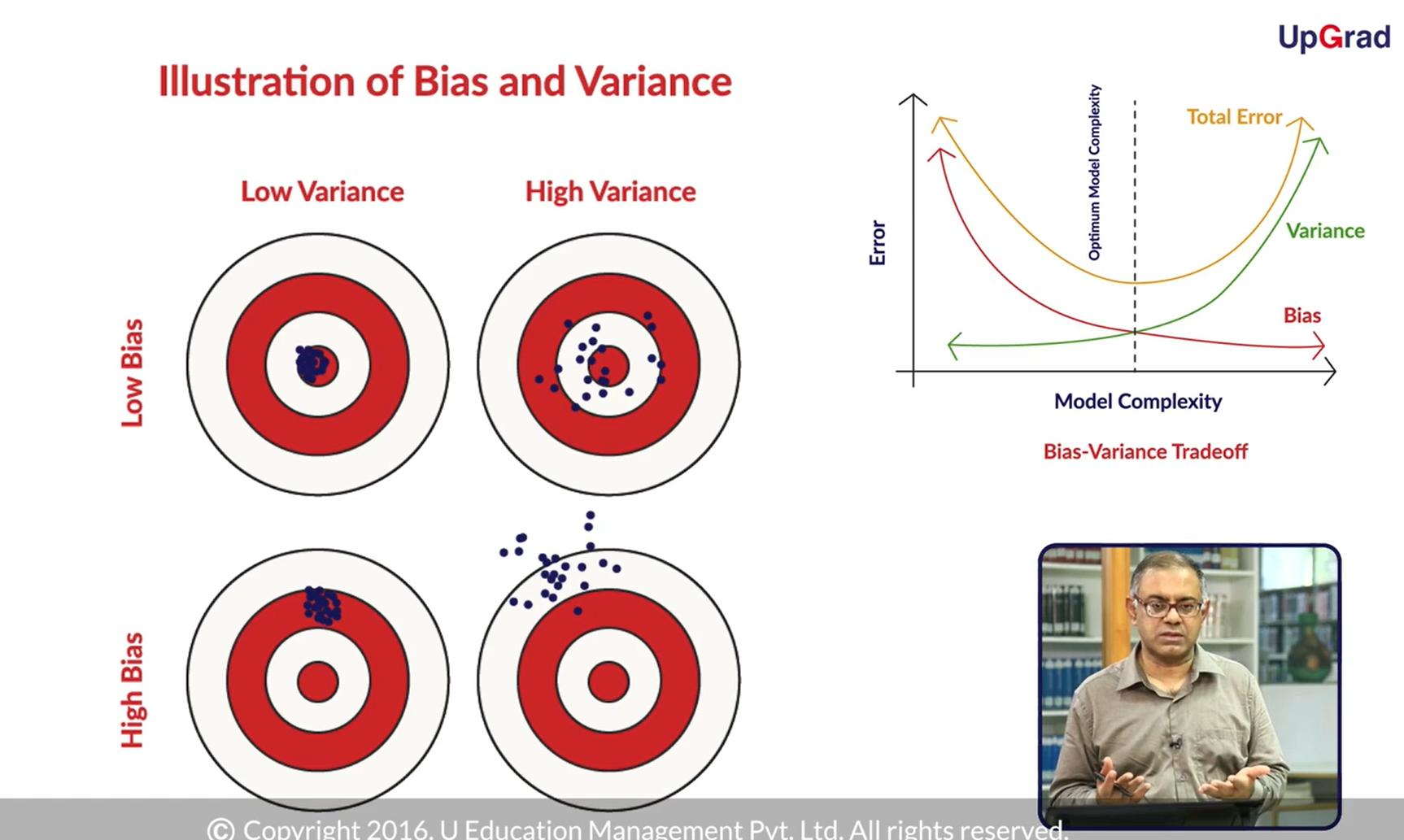
* **Bias-Variance Trade-off.** 

Figure 4/ Bias and Variance: The model should not be very simple so as to give many errors. Also if the model is too complex it will mug up the entire training data. Hence, a balance must be decided between simplicity and complexity of the model.

**Variance**:The ‘variance’ of a model is the **variance in its output** on some test data with respect to the changes in the training data. In other words, here, variance refers to the **degree of changes in the model itself** with respect to changes in the training data.

**Bias:** Bias quantifies how **accurate the model is likely to be**on future (test) data. Extremely simple models are likely to fail in predicting complex real-world phenomena. Simplicity has its own disadvantages.

* **Regularization:**

Regularization is the process of deliberately simplifying models to achieve the correct balance between keeping the model simple and not too naive.