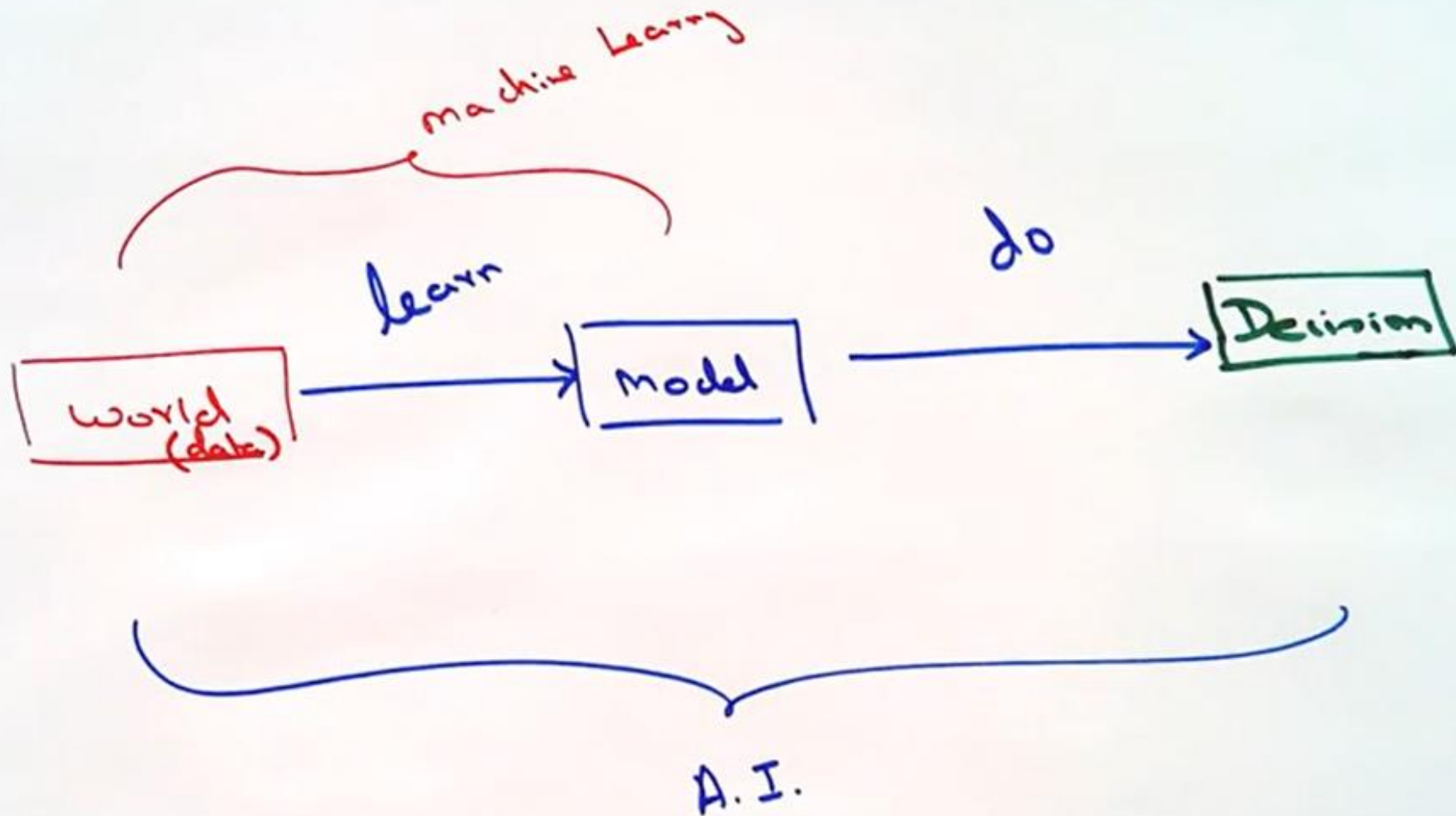


does something  
Smart!

# Intro to Artificial Intelligence and Machine Learning



P	S
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.



Data Sci

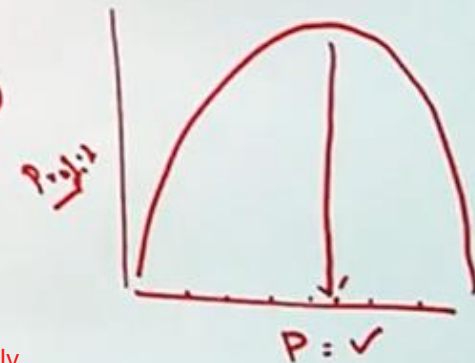
- Statistics
- M.L.

$$\text{Sales} = 1000 - (\text{price}) \times 10$$

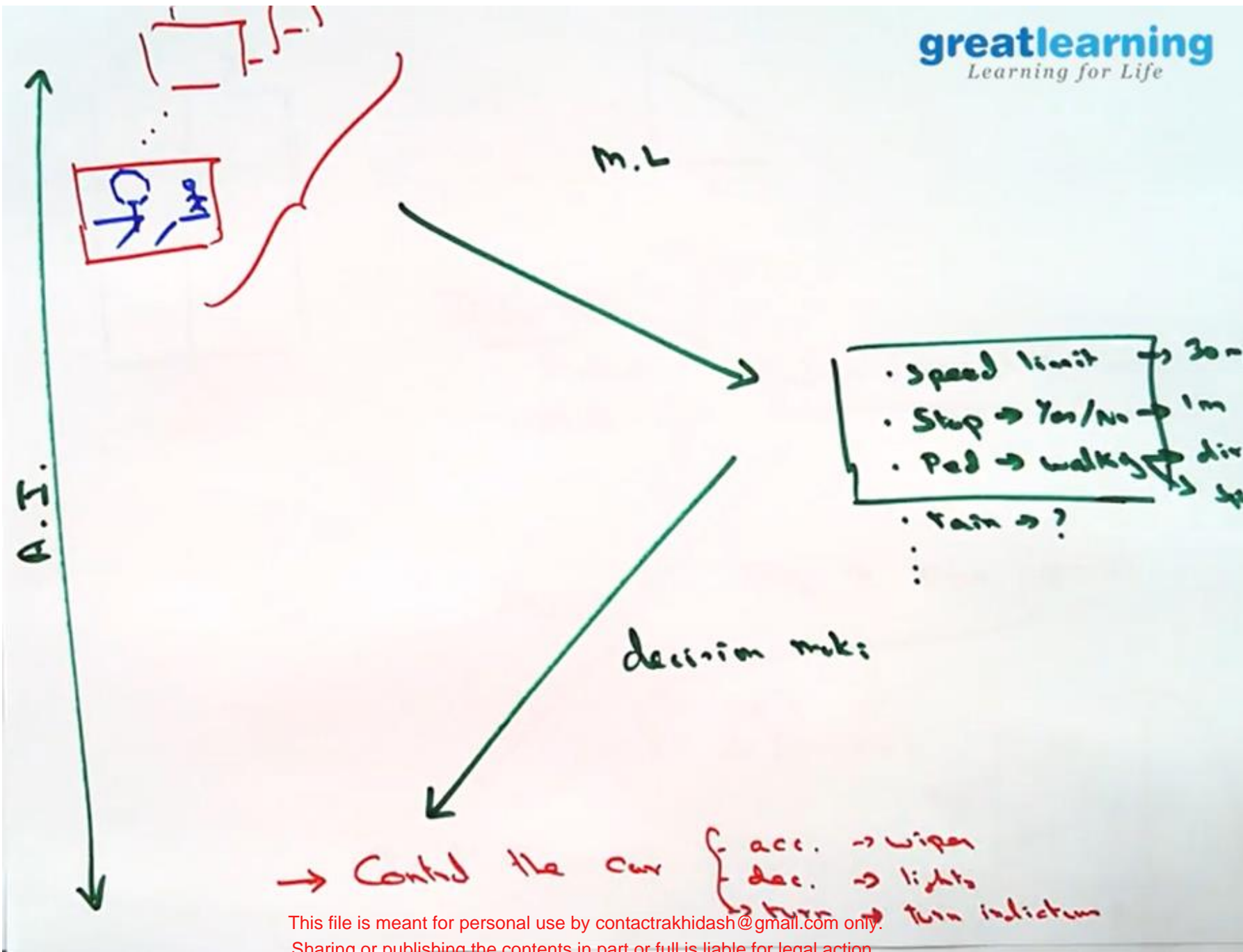
Decision Sci

Obj  $\rightarrow$  max profit

$\hookrightarrow$   $\checkmark$   
 $S(P - \text{cost})$



$P = ? \checkmark$



- Will use Python libraries like,
  - ML with scikit-learn
  - Data handling with Pandas, Numpy, Scipy
  - NLP with NLTK
  - Keras
  - Tensor Flow
  - Seaborn

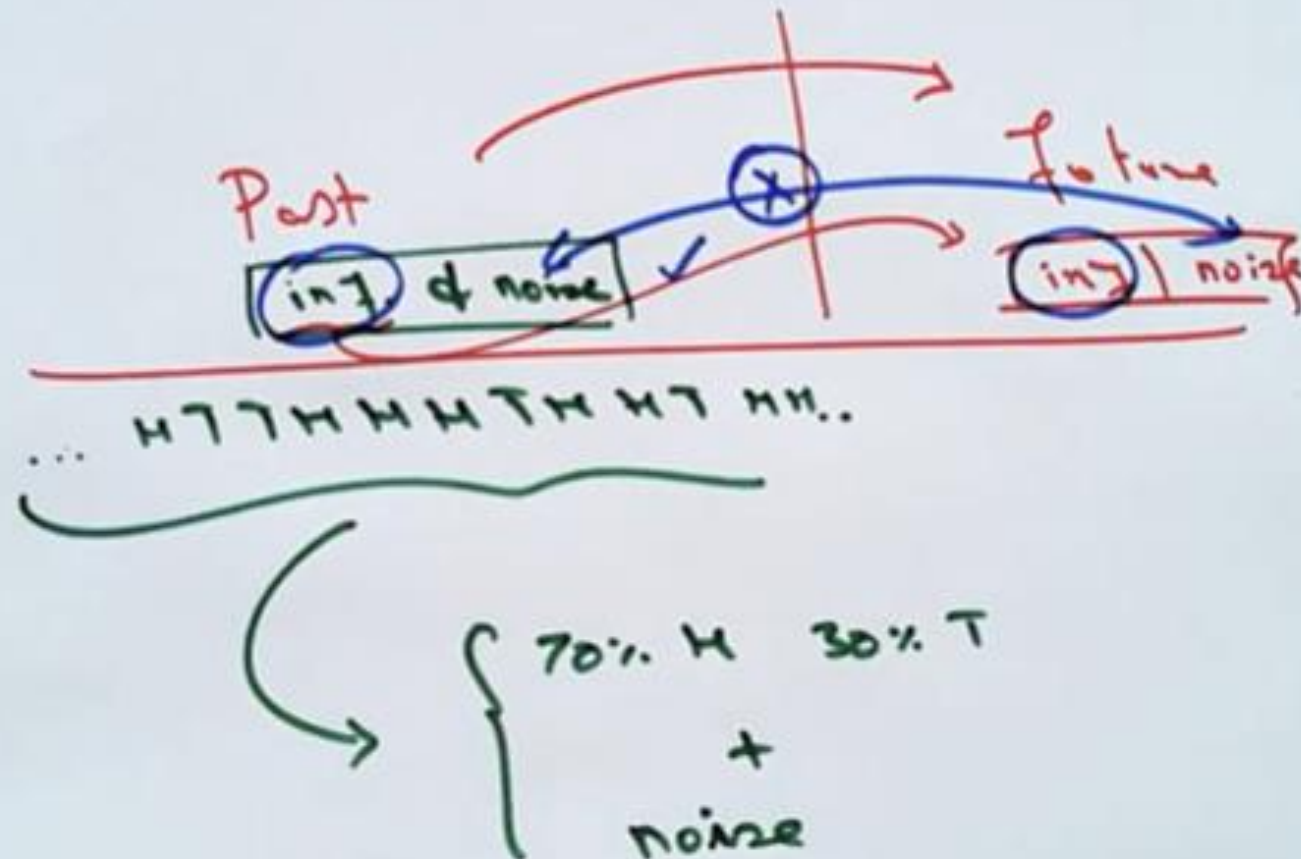


# Program Overview: Machine Learning

- Supervised Learning
- Unsupervised Learning
- Ensemble Techniques
- Recommendation Systems

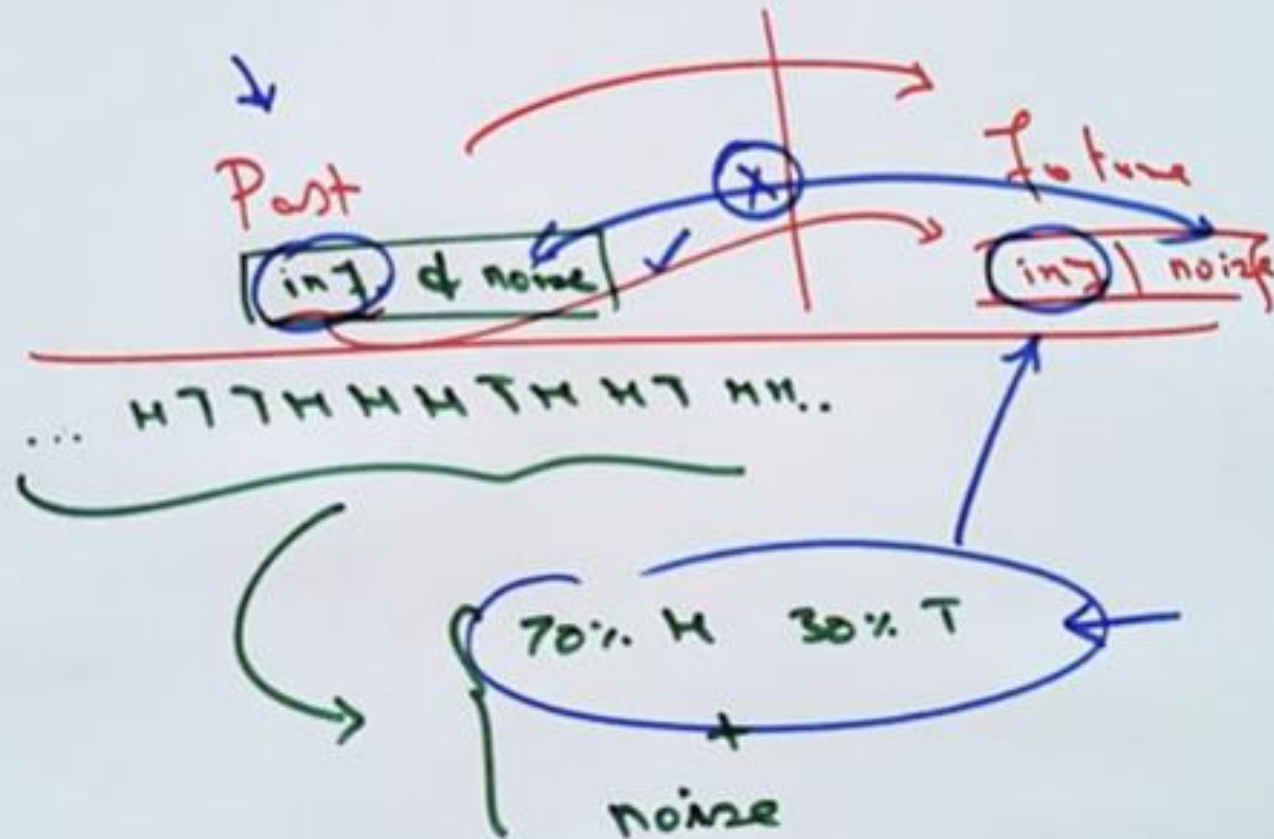
- Neural Network Basics ↩
- Computer Vision ↩
- Statistical NLP
- Sequential NLP
- Advanced Computer Vision

→ learning about the world from data.





→ learning about the world from data



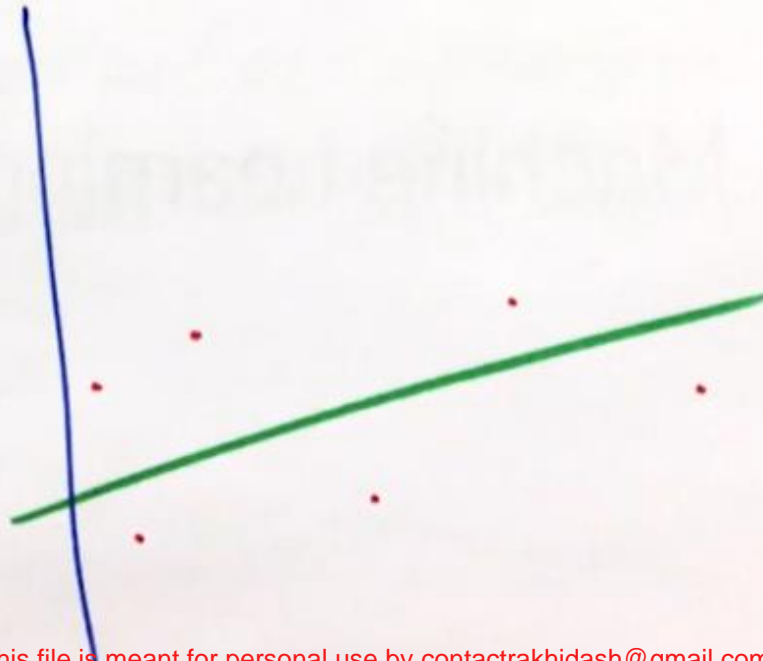
Data

GPA	Annual Sal
:	:
:	:
:	:
:	:

Math model

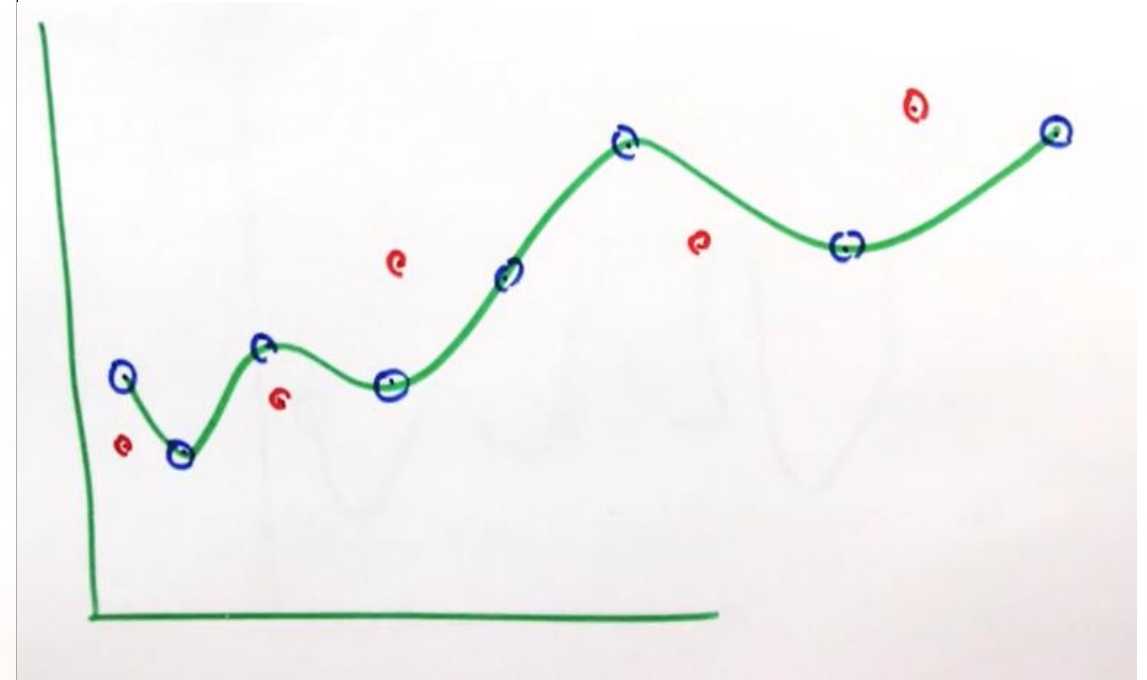
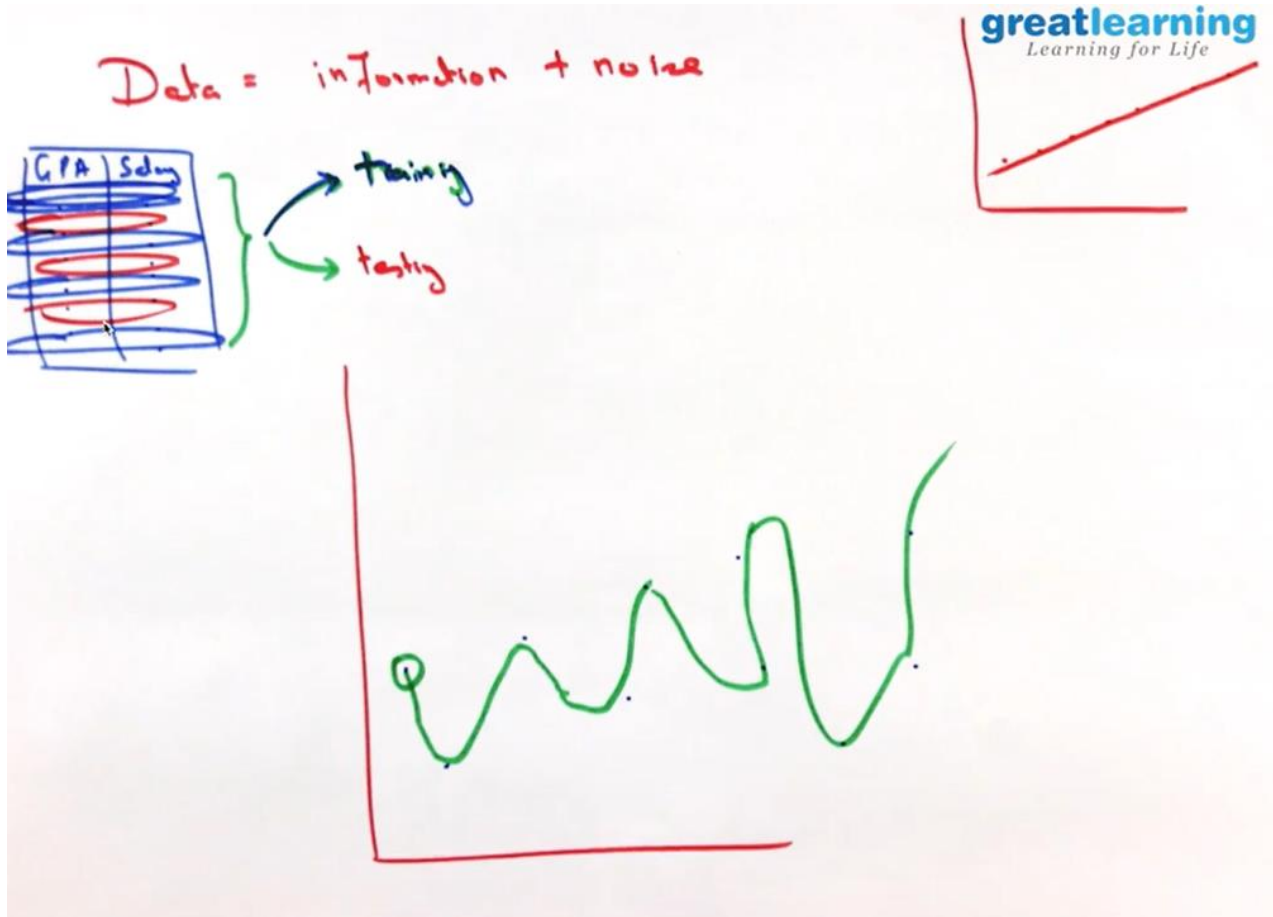
Summary  
 $Salary = a + b(GPA)$

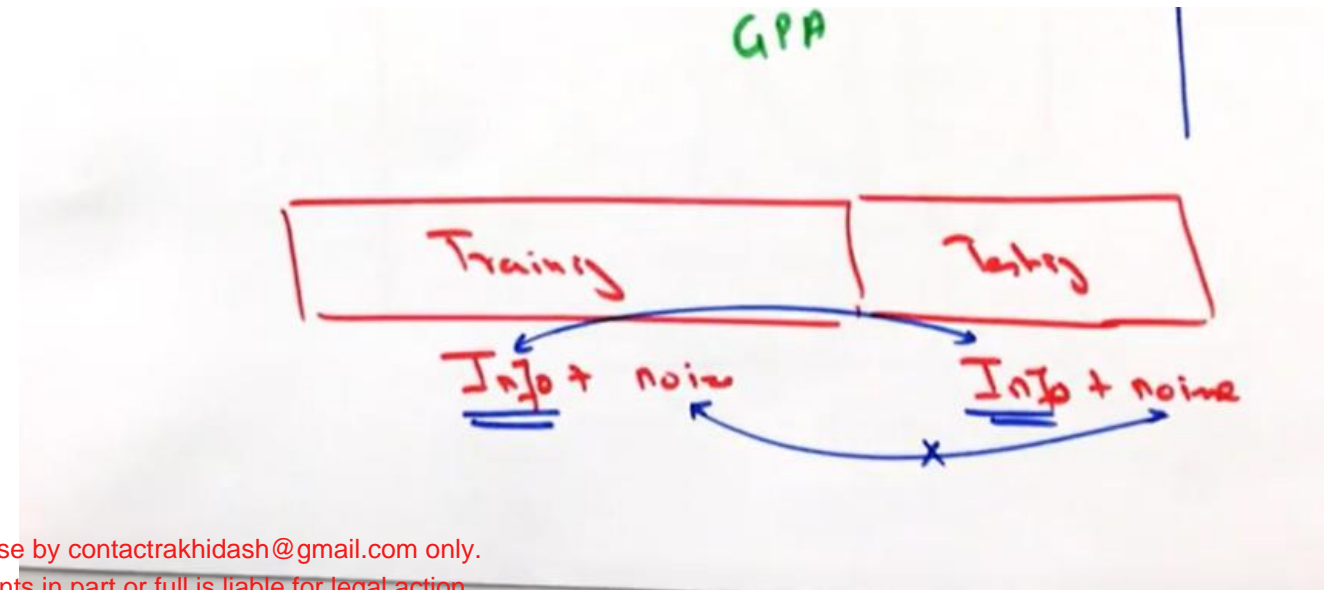
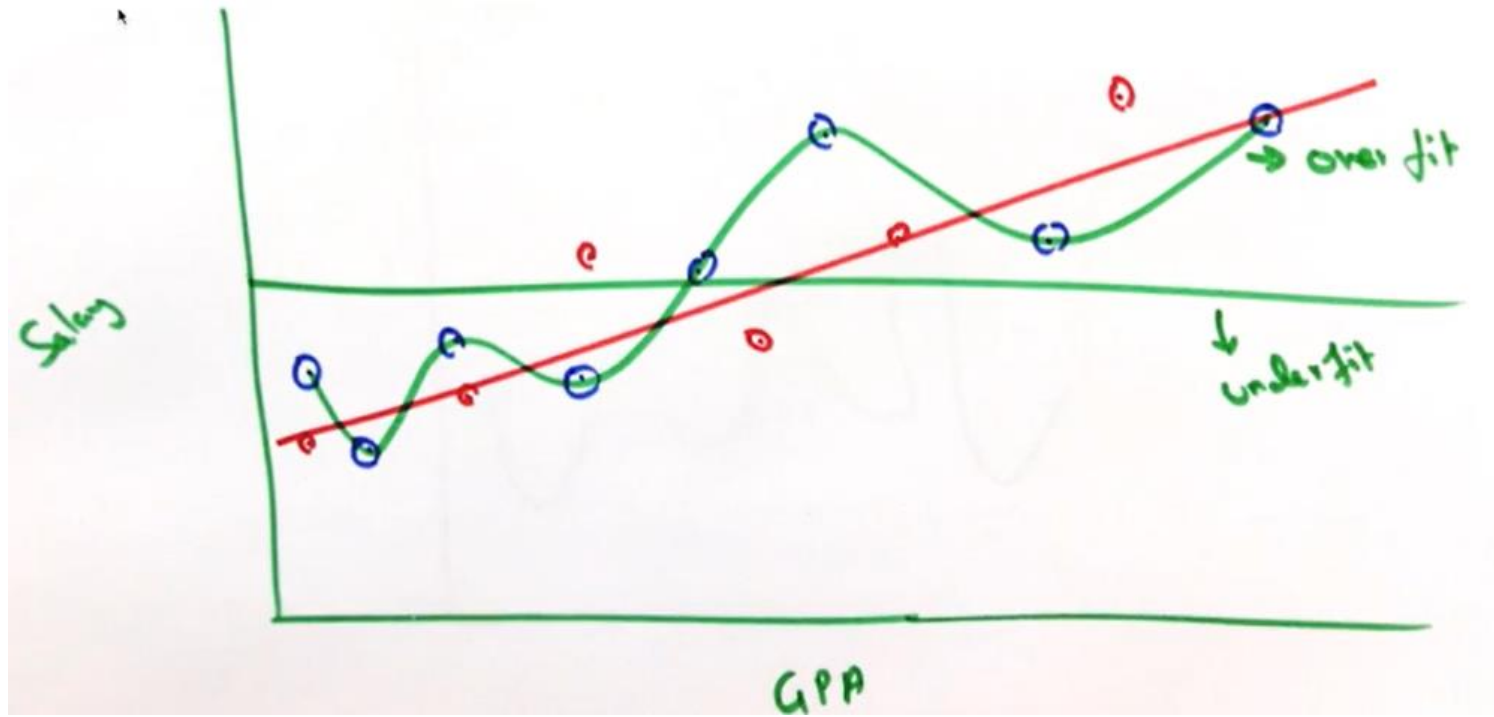
Salary



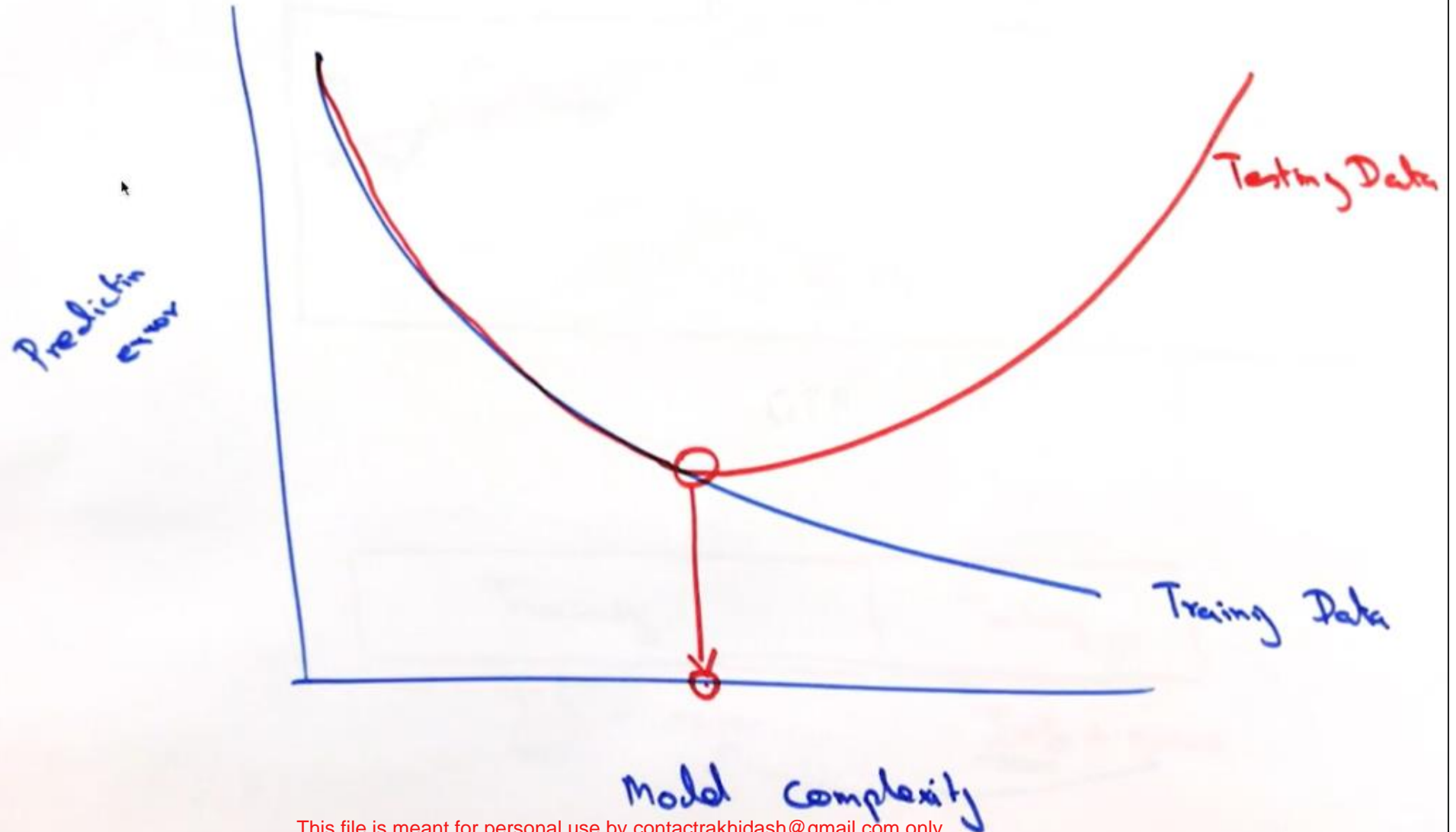
# Learning from Data

- Can we learn about the world around us using data?
- Model building from data
  - Take data as input
  - Find patterns in the data
  - Summarize the pattern in a mathematically precise way
- Machine learning automates this model building.







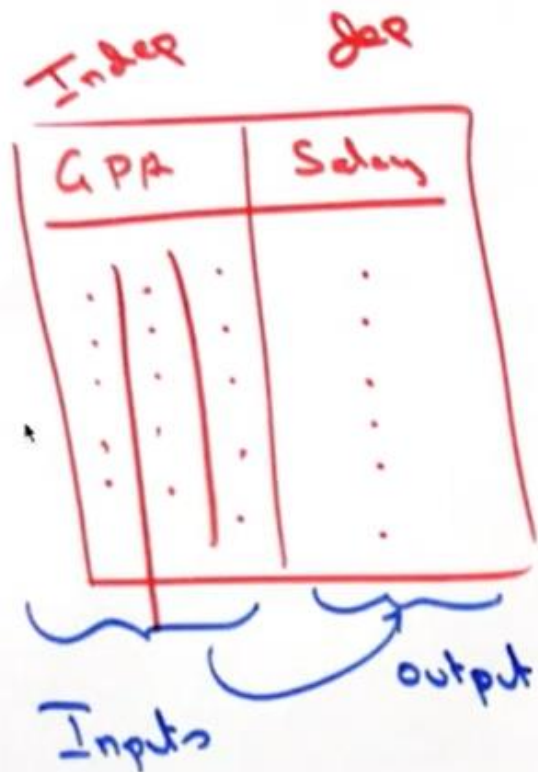


# The Challenge

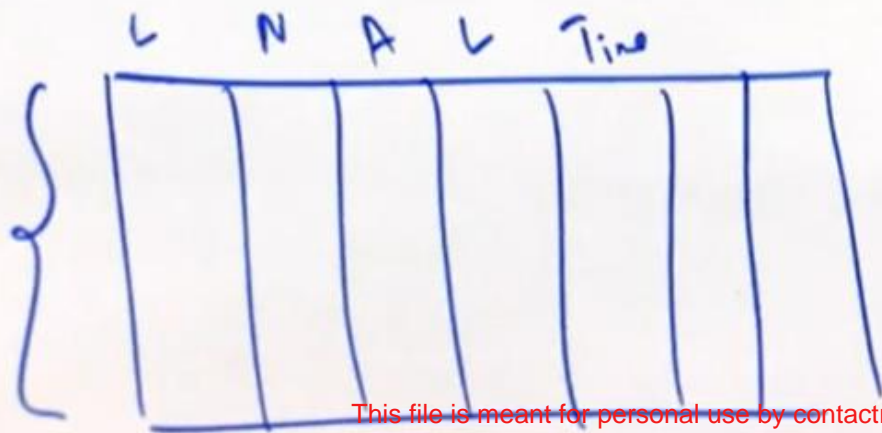
- Data unfortunately contains noise. If not, machine learning would be trivial!
- Think of  $\text{Data} = \text{Information} + \text{Noise}$
- The challenge is to identify the information content and distill away the noise.
- To help do this, machine learning uses a train and test approach.

# Over fitting Vs under fitting

- If the model we finish with ends up
  - modeling the noise as well, we call it “over fitting” - bad for prediction!
  - not modeling all the information, we call it “under fitting” - bad for prediction!
- The hope is that the model that does the best on testing data manages to capture/model all the information but leave out all the noise.



~~Sup~~ Supervised Learning



UnSupervised Learning

# Machine Learning tasks

1. **Supervised learning**: Building a mathematical model using data that contains both the inputs and the desired outputs (ground truth).
  - Examples:
    - Determining if an image has a horse. The data would include images with and without the horse (the input), and for each image we would have a label (the output) indicating if there is a horse in that image.
    - Determining if a client might default on a loan
    - Determining if a call center employee is likely to quit
  - Since we have desired outputs, model performance can be evaluated by comparisons.



# Machine Learning Tasks

2. **Unsupervised learning**: Building a mathematical model using data that contains only inputs and no desired outputs.

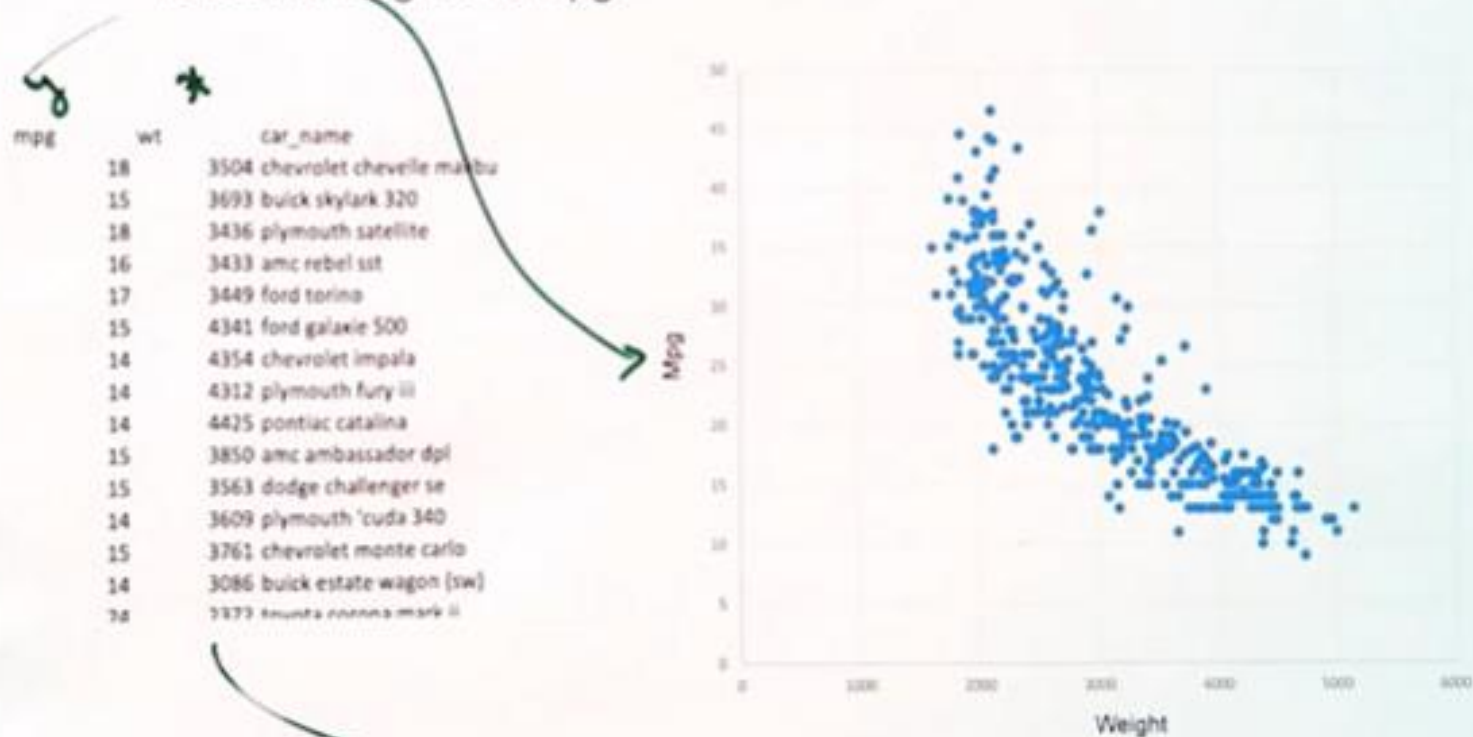
- Used to find structure in the data, like grouping or clustering of data points. To discover patterns and group the inputs into categories.
- Example: an advertising platform segments the population into smaller groups with similar demographics and purchasing habits. Helping advertisers reach their target market with relevant ads.
- Since no labels are provided, there is no specific way to compare model performance in most unsupervised learning methods.

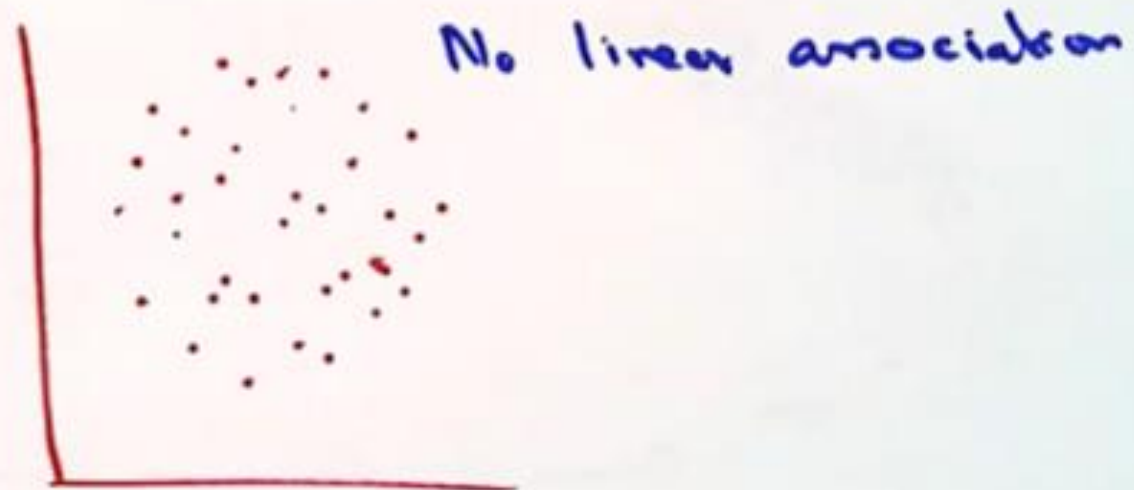
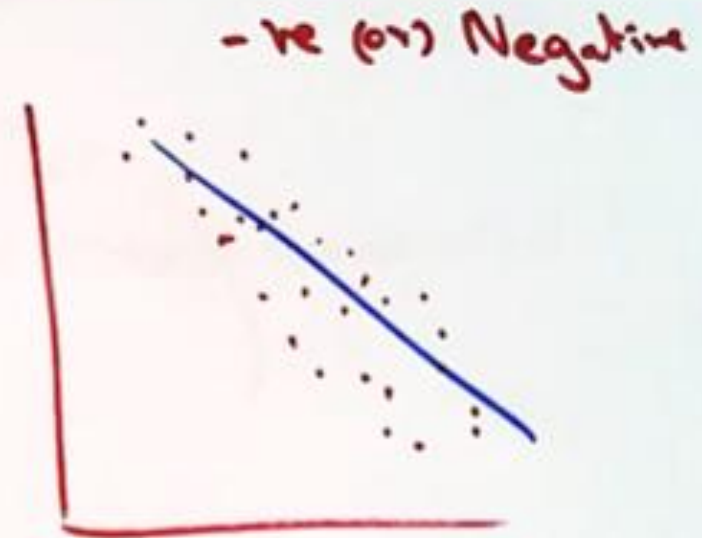
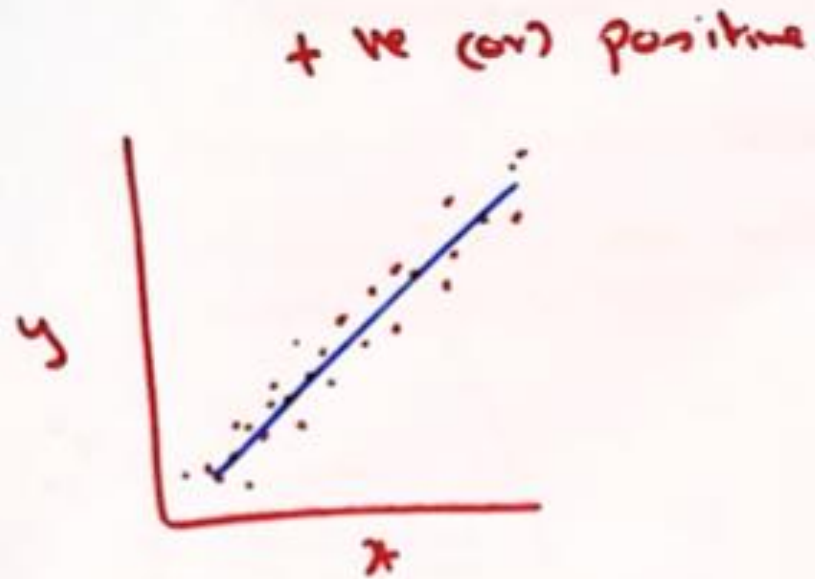
- Supervised learning
  - Regression: desired output is a continuous number
  - Classification: desired output is a category
- Unsupervised learning
  - Clustering: Grouping data
  - Dimensionality reduction: Compressing data
  - Association rule learning: If X then Y



## Linear Relations between two variables

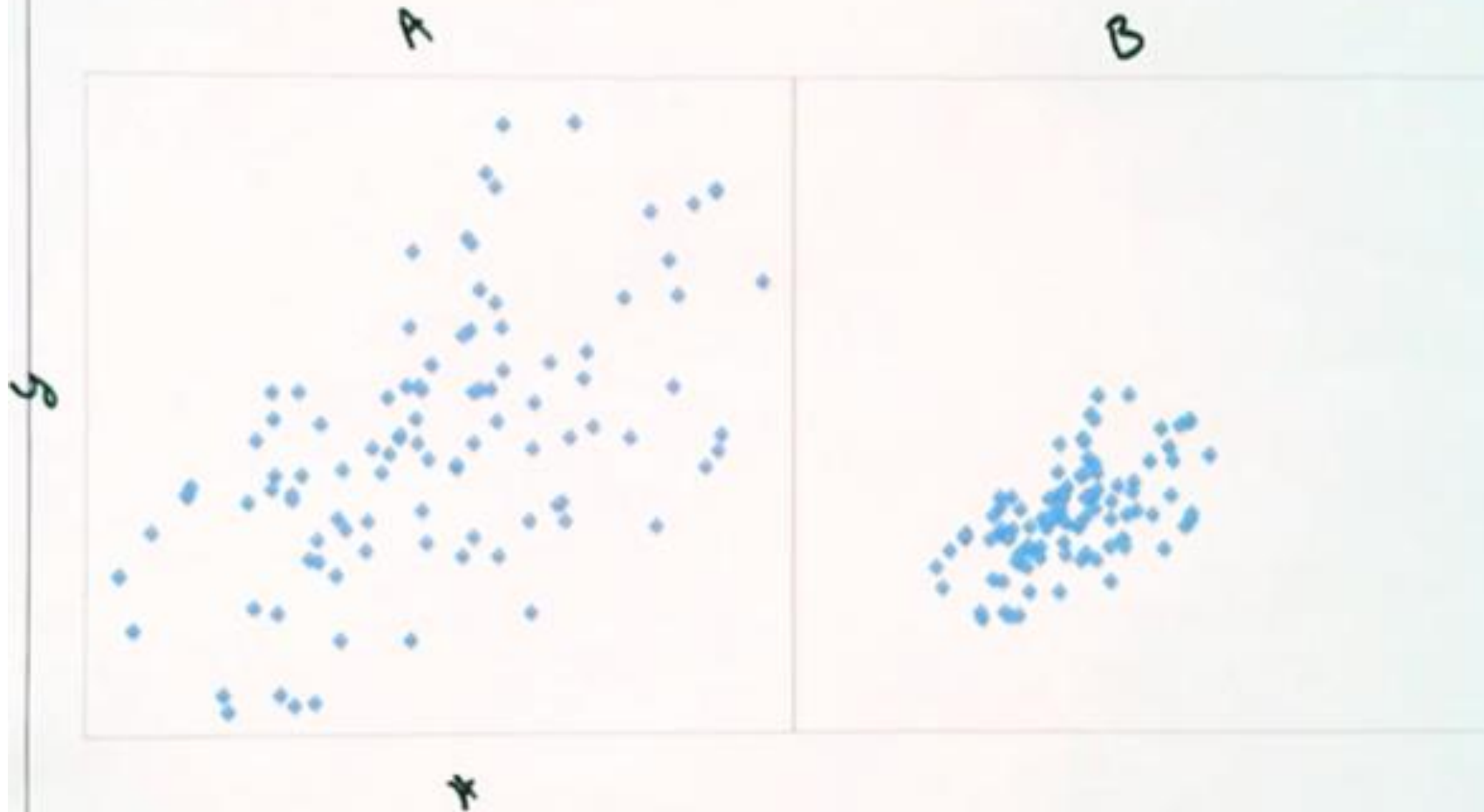
- Do heavier cars have lower mileage?
- Can we use DATA to better understand relationships between the two variables: weight and mpg?







Which one has a stronger relationship?

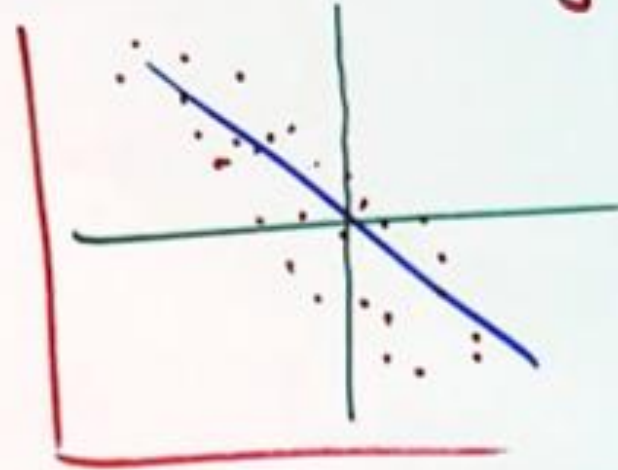




+ve (or) positive

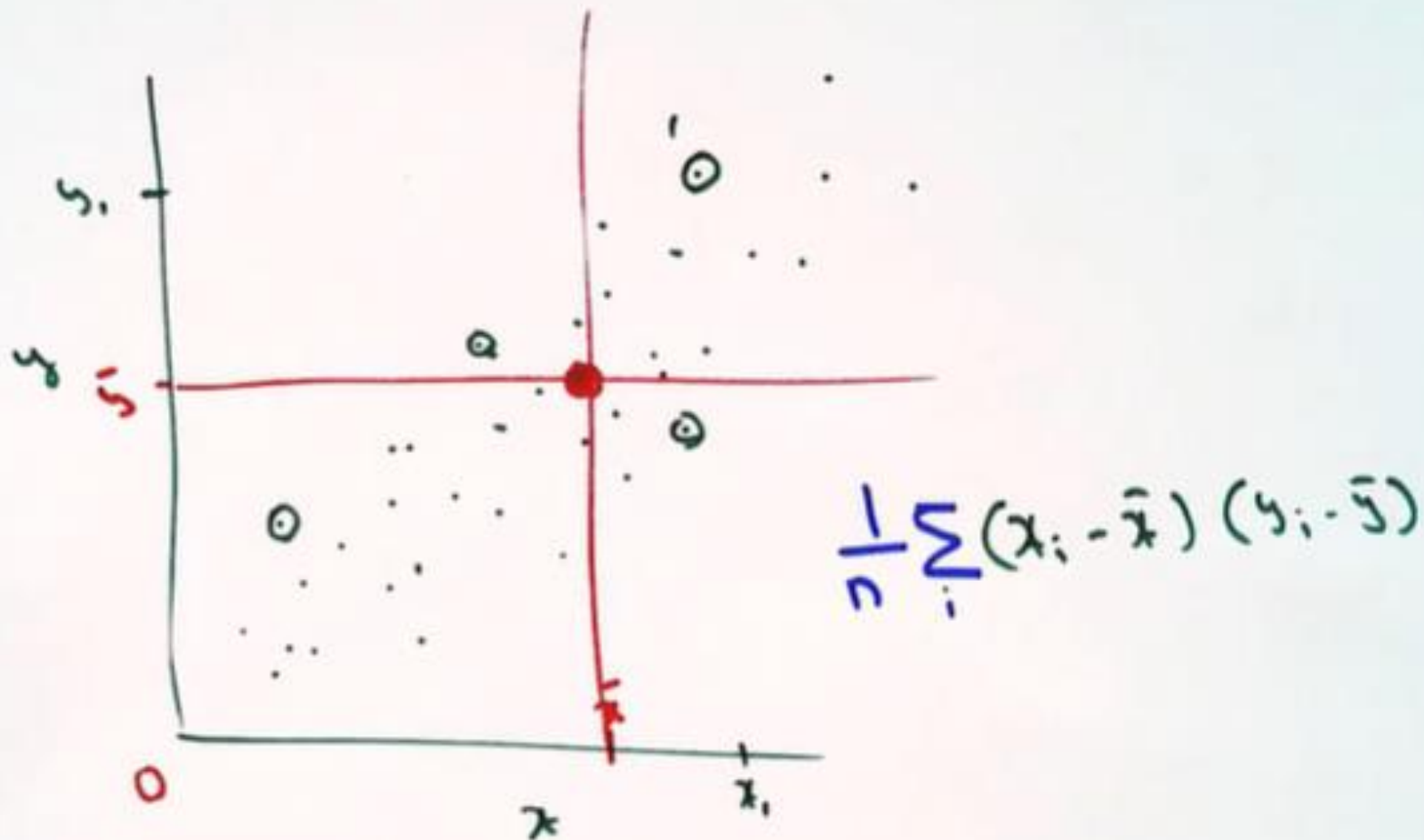


-ve (or) Negative



No linear association





## Measures of Association

- Need a measure of association between two variables.
- By association we mean the strength (and direction) of a linear relationship between two numerical variables.
- The relationship is "strong" if the points in a scatterplot cluster tightly around some straight line. If this line rises from left to right then the relationship is "positive". If it falls from left to right then the relationship is "negative".
- We know that variance of a variable X is

$$Var(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

- On a similar note let's define "covariance" between X and Y as

$$\rightarrow \underline{\underline{Cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}}}$$

$$\text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{\text{Std Dev}(x) \text{ Std Dev}(y)} \Rightarrow \text{does not have any unit}$$

$\downarrow$  (unit x)       $\downarrow$  (unit y)

$$\text{Corr}(x, x) = 1$$

$$\text{Corr}(x, -x) = \frac{-\text{Var}(x)}{\text{Var}(x)} = -1$$



- Covariance:
  - Covariance between X and Y is the same as the covariance between Y and X.
  - The covariance between a variable and itself is the variance of the variable.
  - It is difficult to interpret the magnitudes of covariances since it is not scale invariant.

- Correlation
  - We can scale covariance to make it an invariant measure of linear association!
  - Correlation between X and Y is

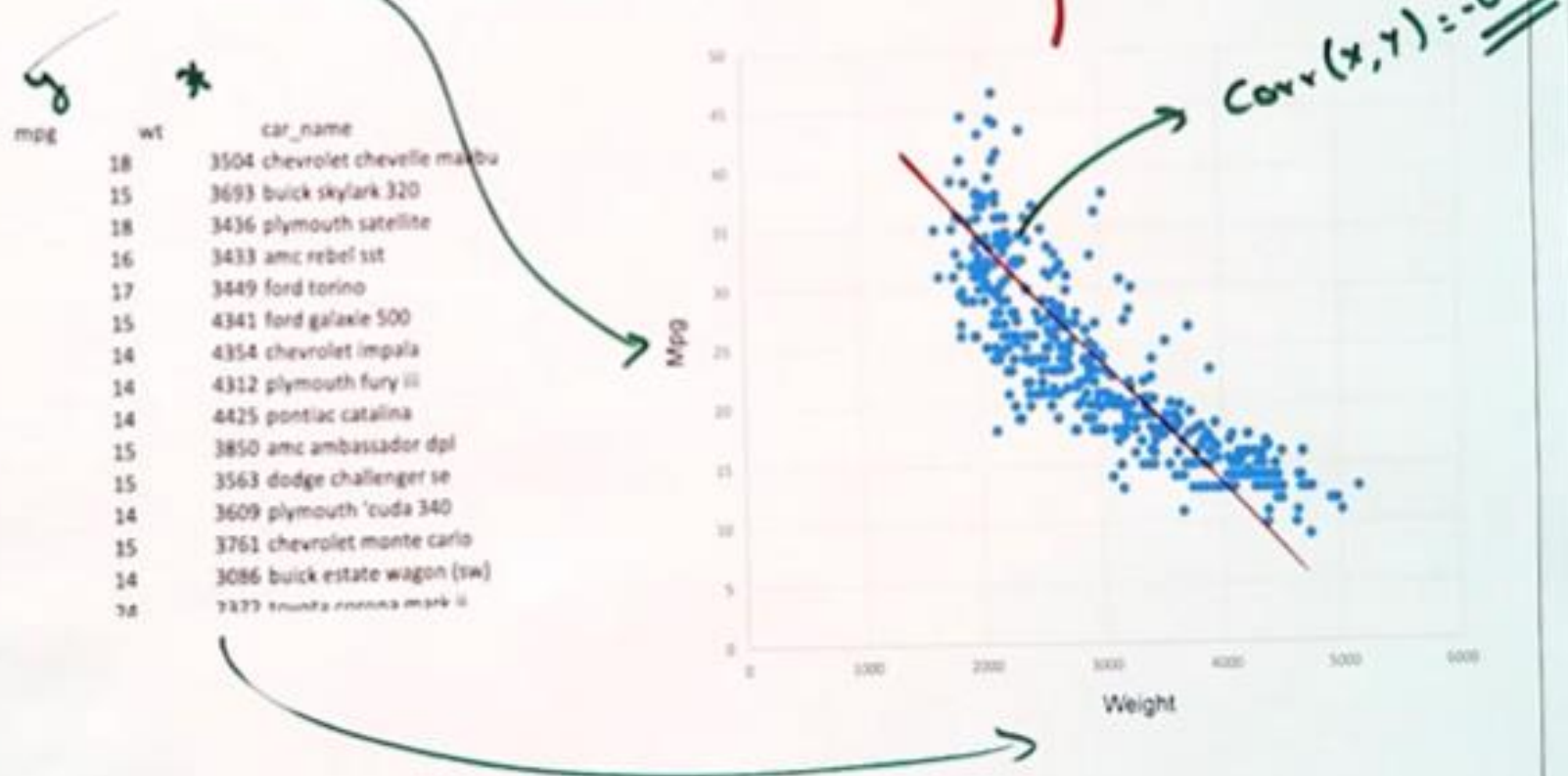
$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{Stdev}(X) \times \text{Stdev}(Y)}$$

- Correlation is always between -1 and +1. The correlation between a variable and itself is 1.
- The correlation between X and Y is the same as the correlation between Y and X.
- Correlation is scale invariant



## Linear Relations between two variables

- Do heavier cars have lower mileage? **Yes**
- Can we use DATA to better understand relationships between the two variables: weight and mpg?



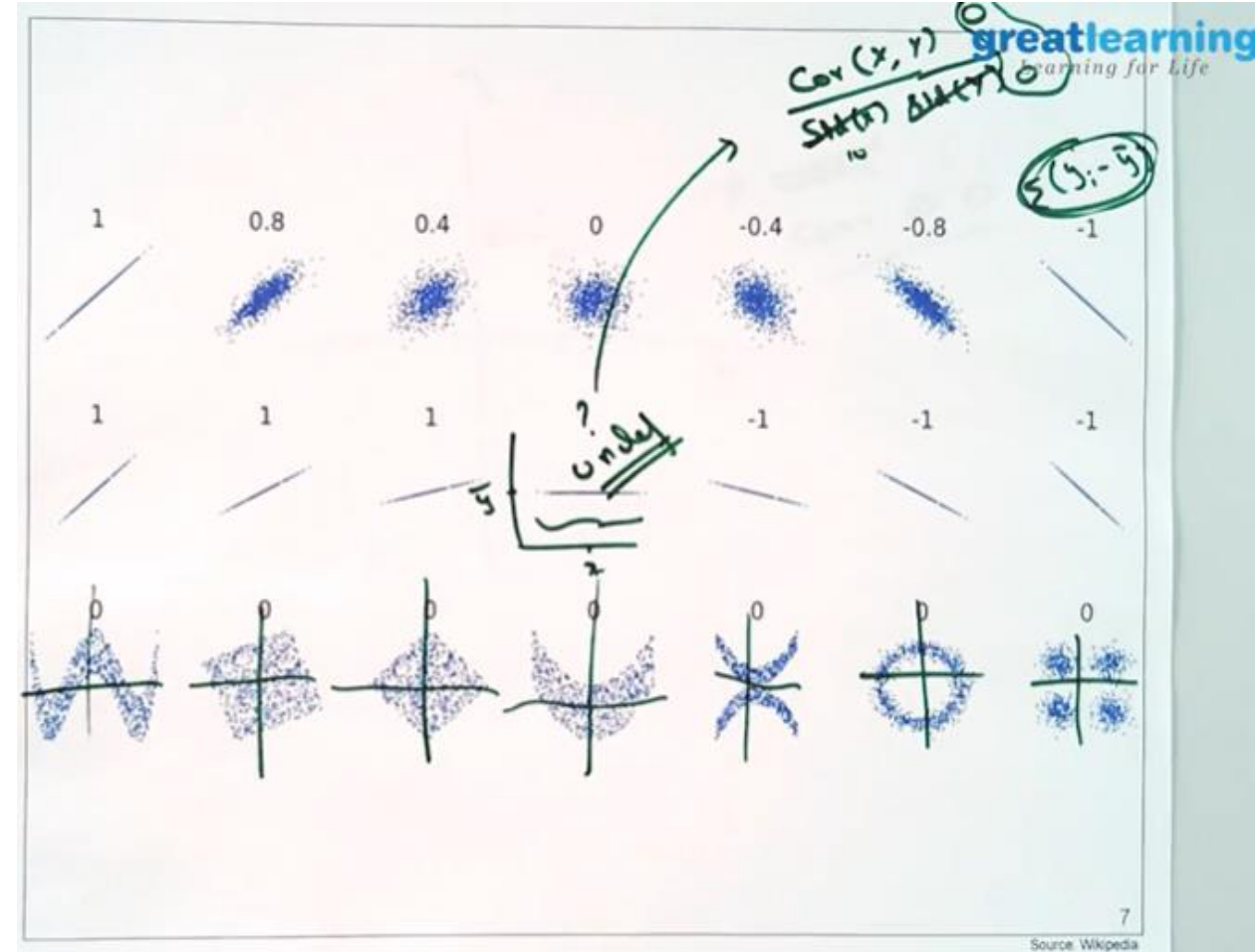
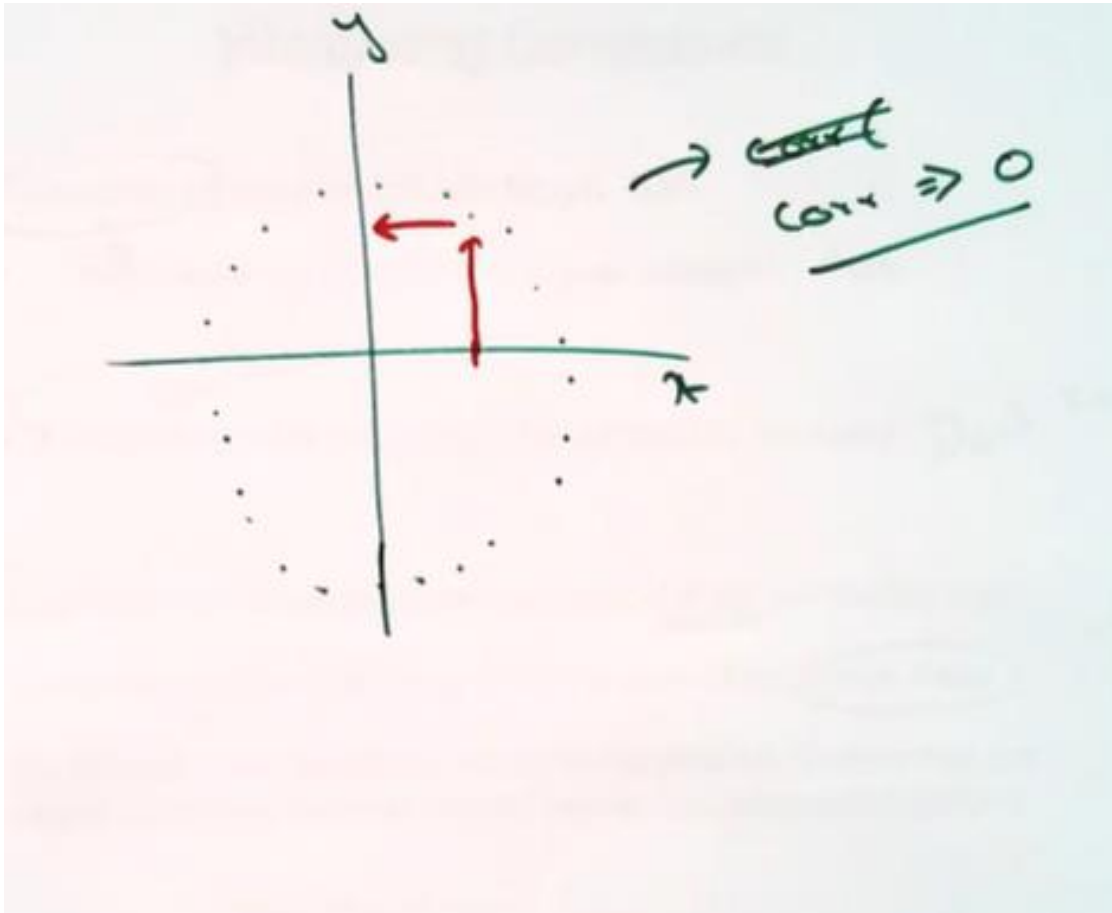
## Interpreting Correlations

- Correlation between Weight and Mpg is -0.83

- Does heavier car tend to have a lower mileage? *Yes*

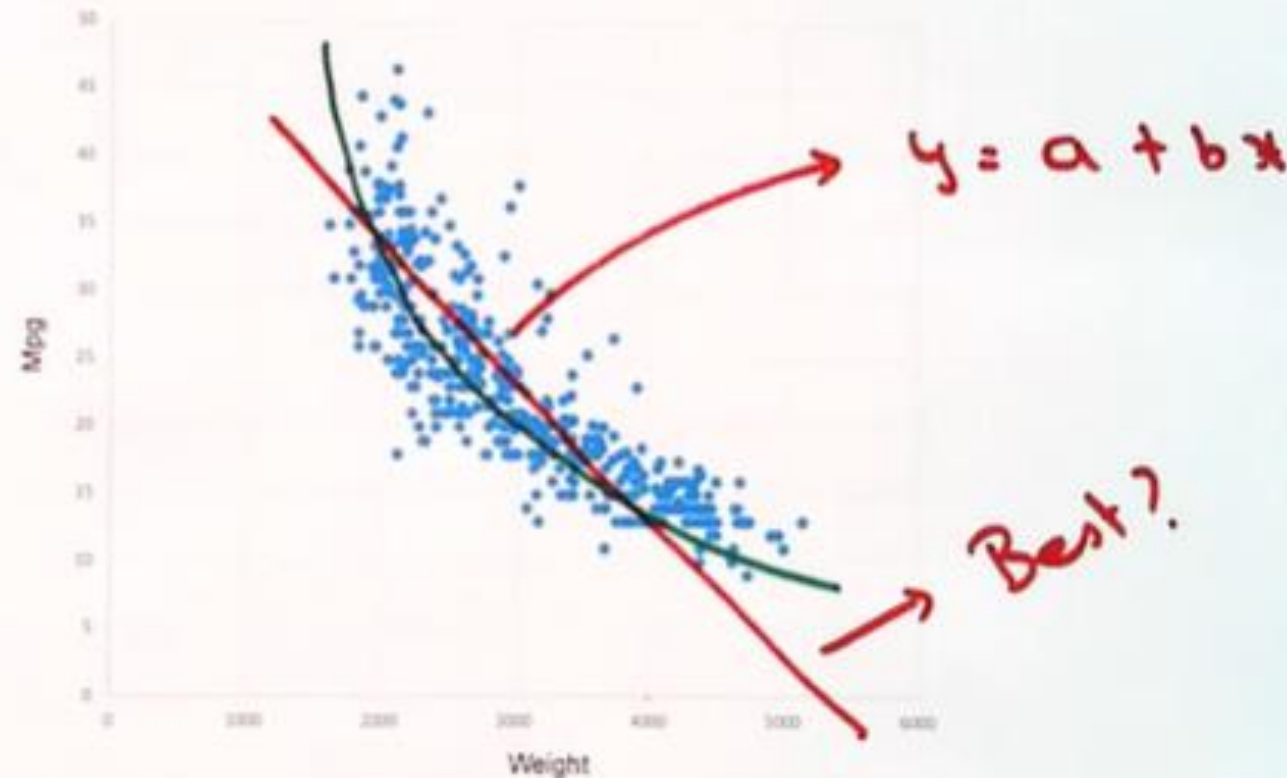
- *Causation*  
→ If we increase the weight of a car, will its Mpg decrease? *Don't Know*

- Correlation and covariance are measures of linear association only.
- Correlation can be misleading when the association is non-linear
- Outliers can have significant effects on correlations. Outliers that are clearly identifiable are best deleted before correlation computations.

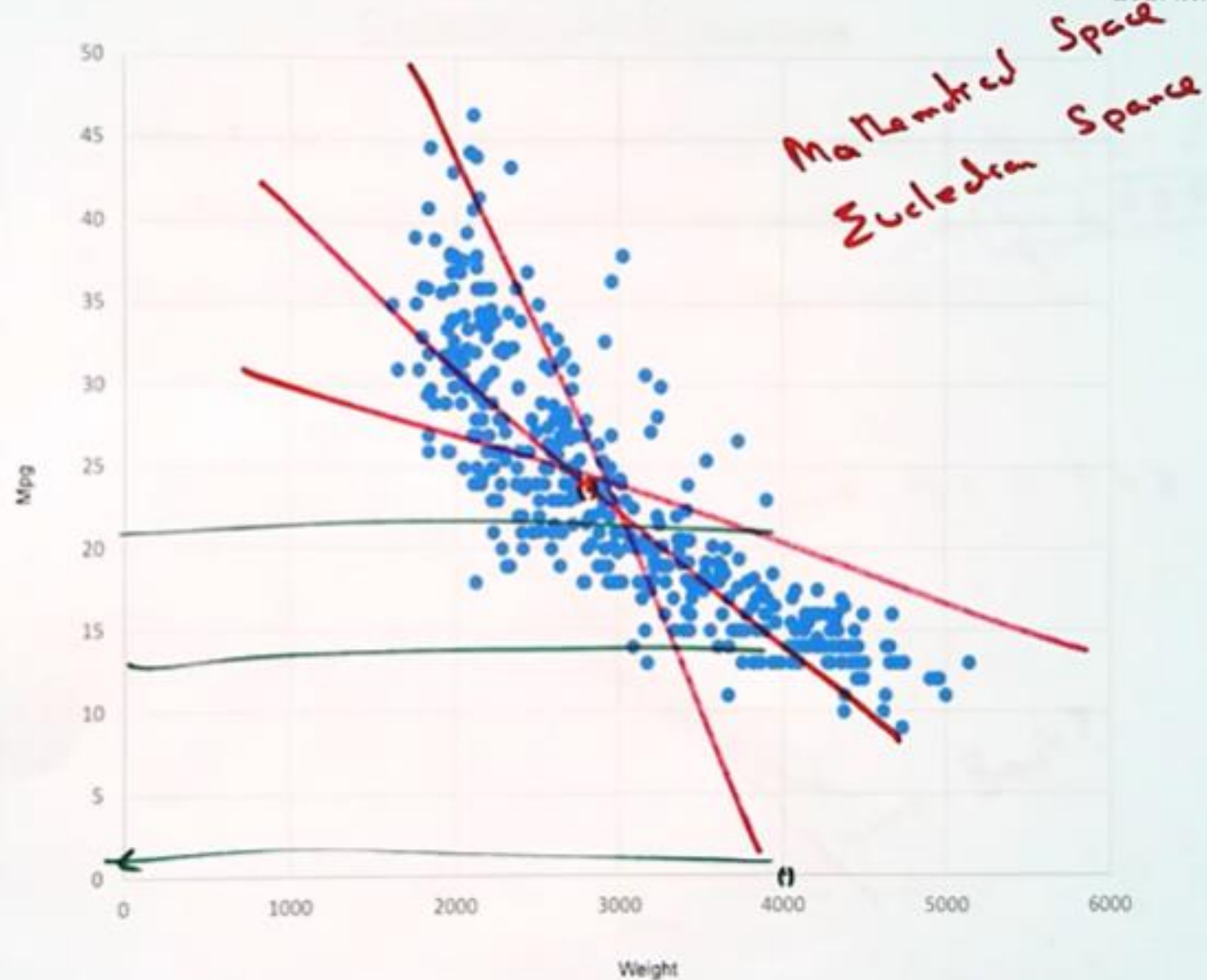


- Next: If a car's weight is 4000, what would we expect its Mpg to be?
- Previously: Measuring strength of relationship
- Now: Capturing relationships using a simple model (equation)

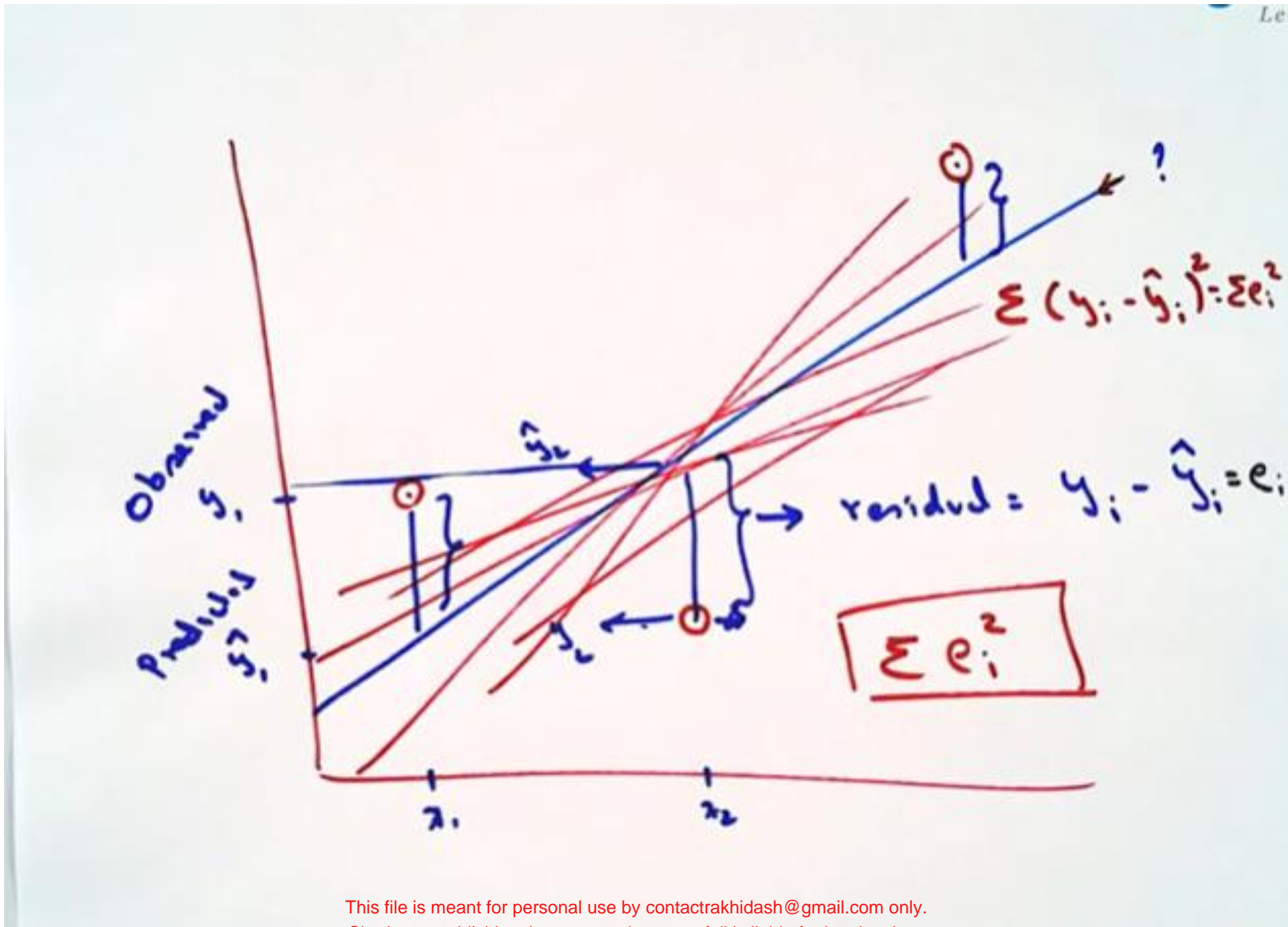
$$y = a + bx$$



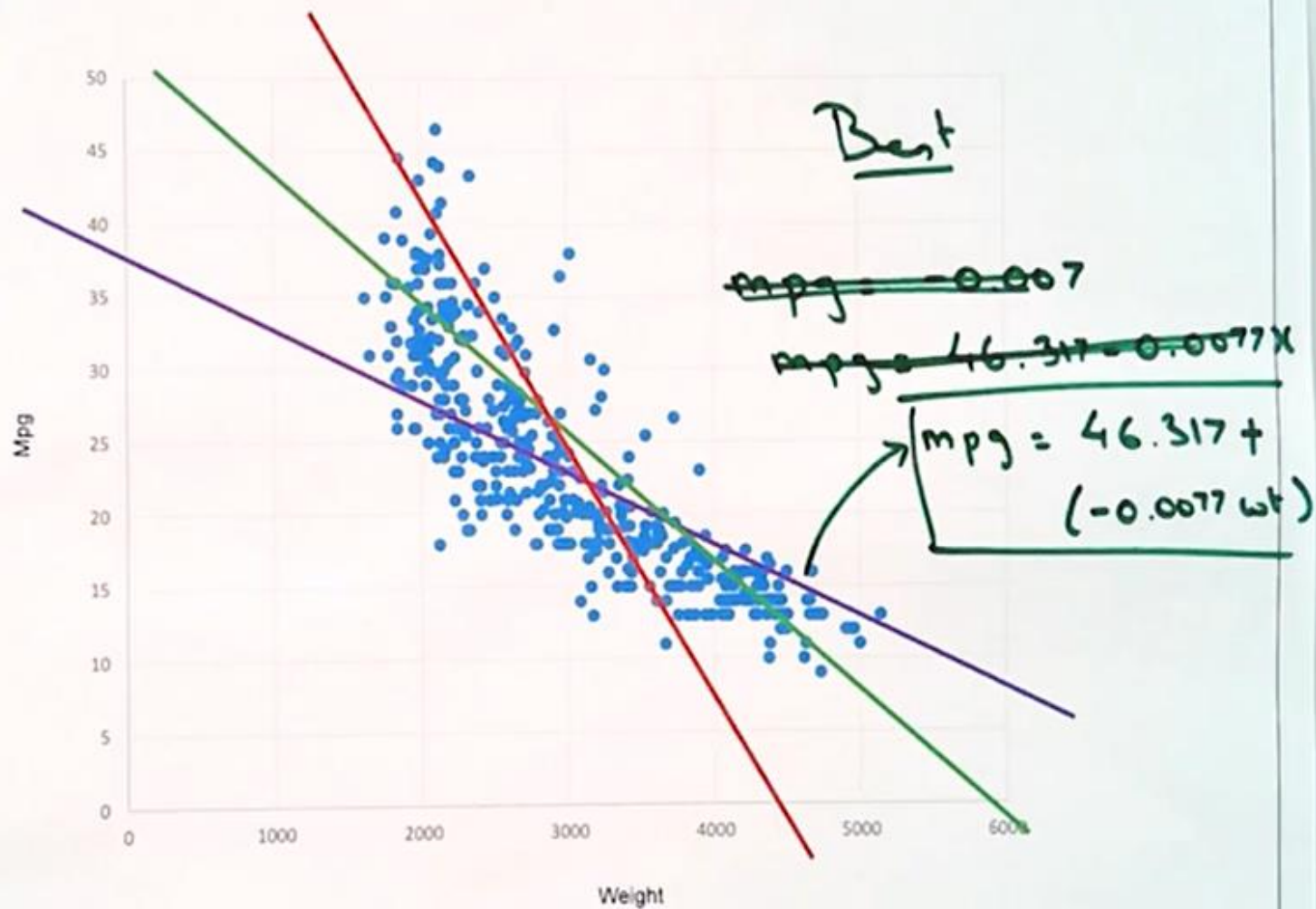








## How easy is it to fit a straight line?



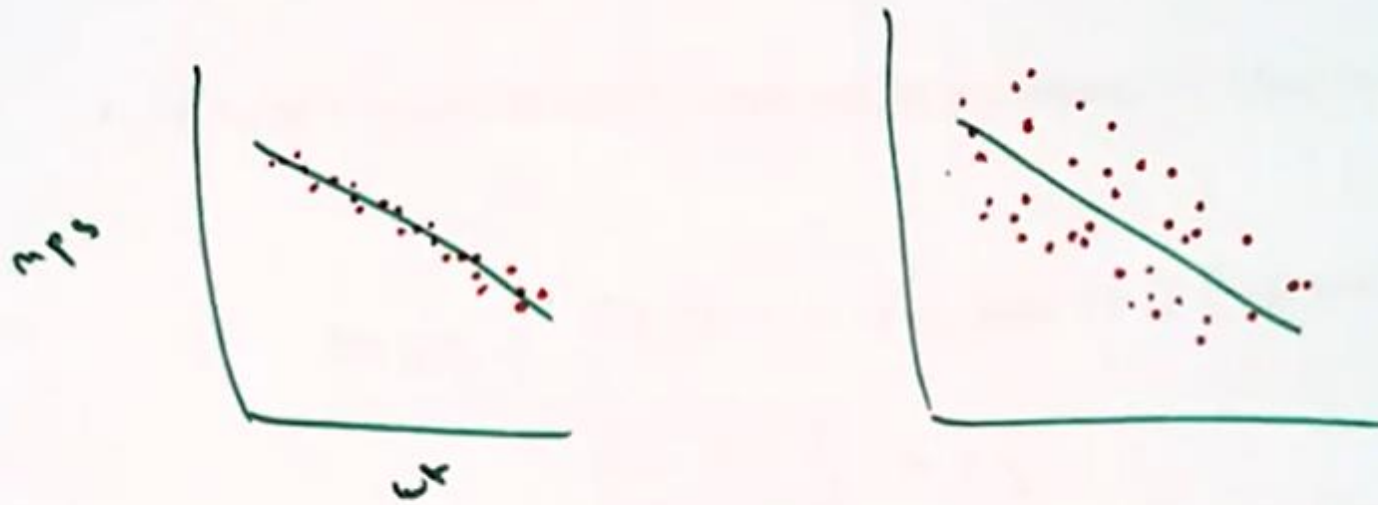
So...

- If a car's weight is 4000, what would we expect its Mpg to be?

$$\begin{aligned} \text{mpg} &= 46.517 - (0.0077)(4000) \\ &= \boxed{15.52} \text{ mpg} \end{aligned}$$

$$\boxed{y = a + bx} \rightarrow \text{linear}$$

- We managed to use the data to construct a regression model. Using this model we answered the above question.





How good is our regression fit?

→ Standard of residuals = Std. error  
of the  
Regression  
 $\sigma_e$





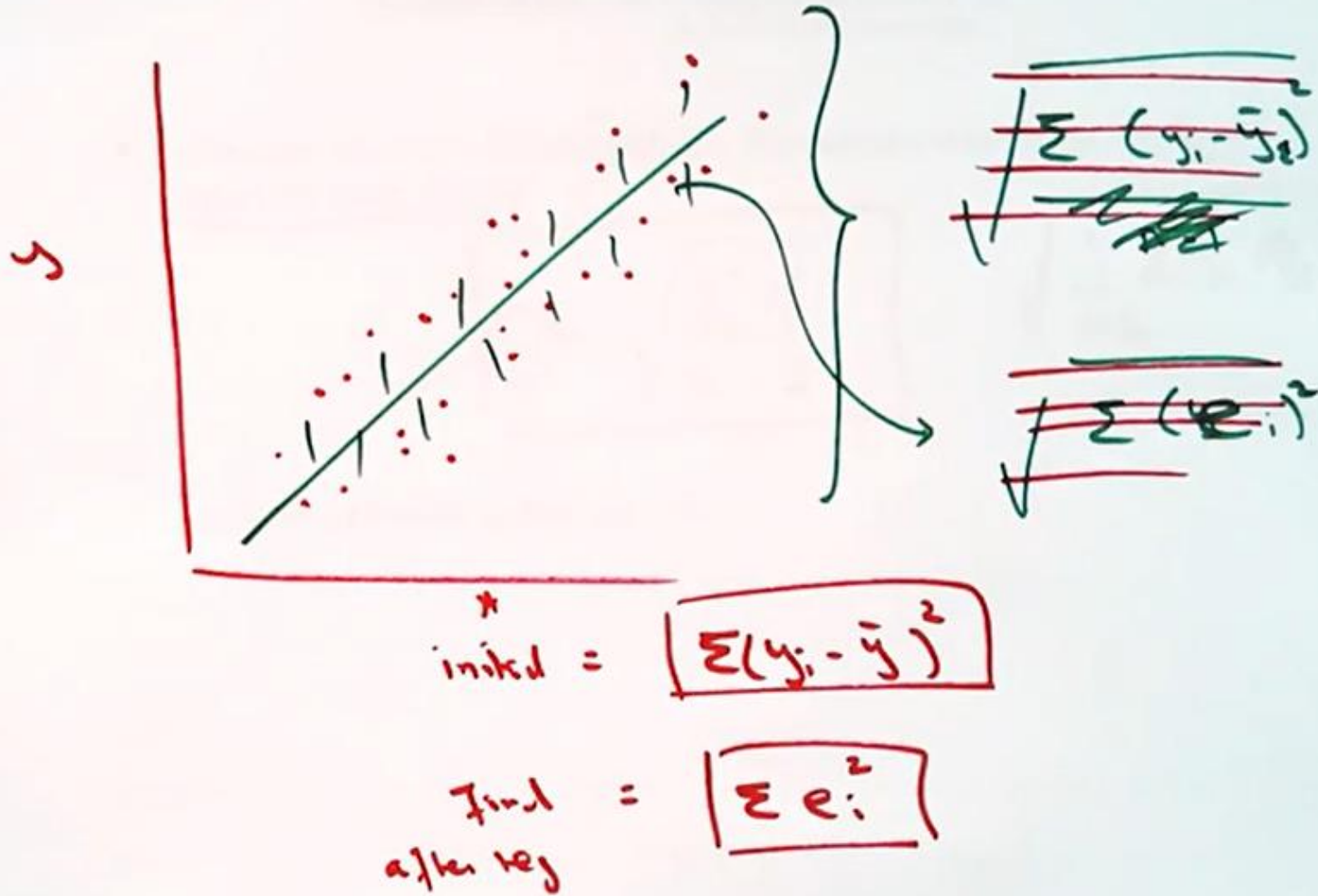
## Measures of Regression Fit

- Standard deviation of the residuals. Sometimes also called the Root Mean Sq Error (RMSE)

$$s_e = \sqrt{\frac{\sum e_i^2}{n-2}}$$

$$\sqrt{\frac{1}{n-2} \sum (y_i - \hat{y}_i)^2}$$

- Comparing RMSE to Std. dev of y



$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}$$

$\sum e_i^2 \rightarrow SSE$   
 $\sum (y_i - \bar{y})^2 \rightarrow SST$

1 Perfect expl.  
 0 expl. nothing

True of the var. in  $y$   
Expl by the Reg.

only in Single  
linear Reg.

$$\Rightarrow (Corr)^2 = R^2$$

$y \rightarrow Corr$

$$R^2 = \frac{SST - SSE}{SST} = \frac{SSR}{SST}$$

# Multiple Regression

- One dependent variable. More than one independent variable.
- The regression model (equation)

$$y = a + b_1x_1 + \dots + b_kx_k \rightarrow \text{Hyper plane}$$

- The above is the equation of a hyper-plane set in k dimensions
- Again use the similar arguments to find the best hyper-plane by minimizing the least squares measure.

- Very easily computed using most Statics of ML tools

mpg	cyl	displacement	hp	wt	acc	yr	origin
18	8	307	130	3504	12	70	American
15	8	350	165	3693	11.5	70	American
18	8	318	150	3436	11	70	American
16	8	304	150	3433	12	70	American
17	8	302	140	3449	10.5	70	American
15	8	429	198	4341	10	70	American
14	8	454	220	4354	9	70	American
14	8	440	215	4312	8.5	70	American
14	8	455	225	4425	10	70	American
15	8	390	190	3850	8.5	70	American

1. mpg: miles per gallon  
2. cyl: cylinders  
3. disp: displacement (cu. inches)  
4. hp: horsepower  
5. wt: weight (lbs)  
6. acc: acceleration (secs for 0-60mph)  
7. yr: model year  
8. origin (American, European, Japanese)  
9. car name



## Standard Error and Adjusted R<sup>2</sup>

- Standard Error for Multiple regression

$$\rightarrow s_e = \sqrt{\frac{\sum e_i^2}{n - k - 1}}$$

Handwritten notes:

- $\sum e_i^2$  (crossed out)
- $\sum (y_i - \bar{y})^2$
- $R^2$  (circled)

- Adjusted R<sup>2</sup>
  - A measure that adjusts for the number of independent variables used
  - Used to monitor if more independent variables belong to the model
  - Cannot be interpreted as "percentage or variation explained"



$$y = a + b_1 x_1$$
$$y = a + b_1 x_1 + b_2 x_2$$
$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3$$

$R^2 = 60\%$   
 $R^2 = \underline{70\%}$   
 $R^2 = \underline{\underline{70.001\%}}$

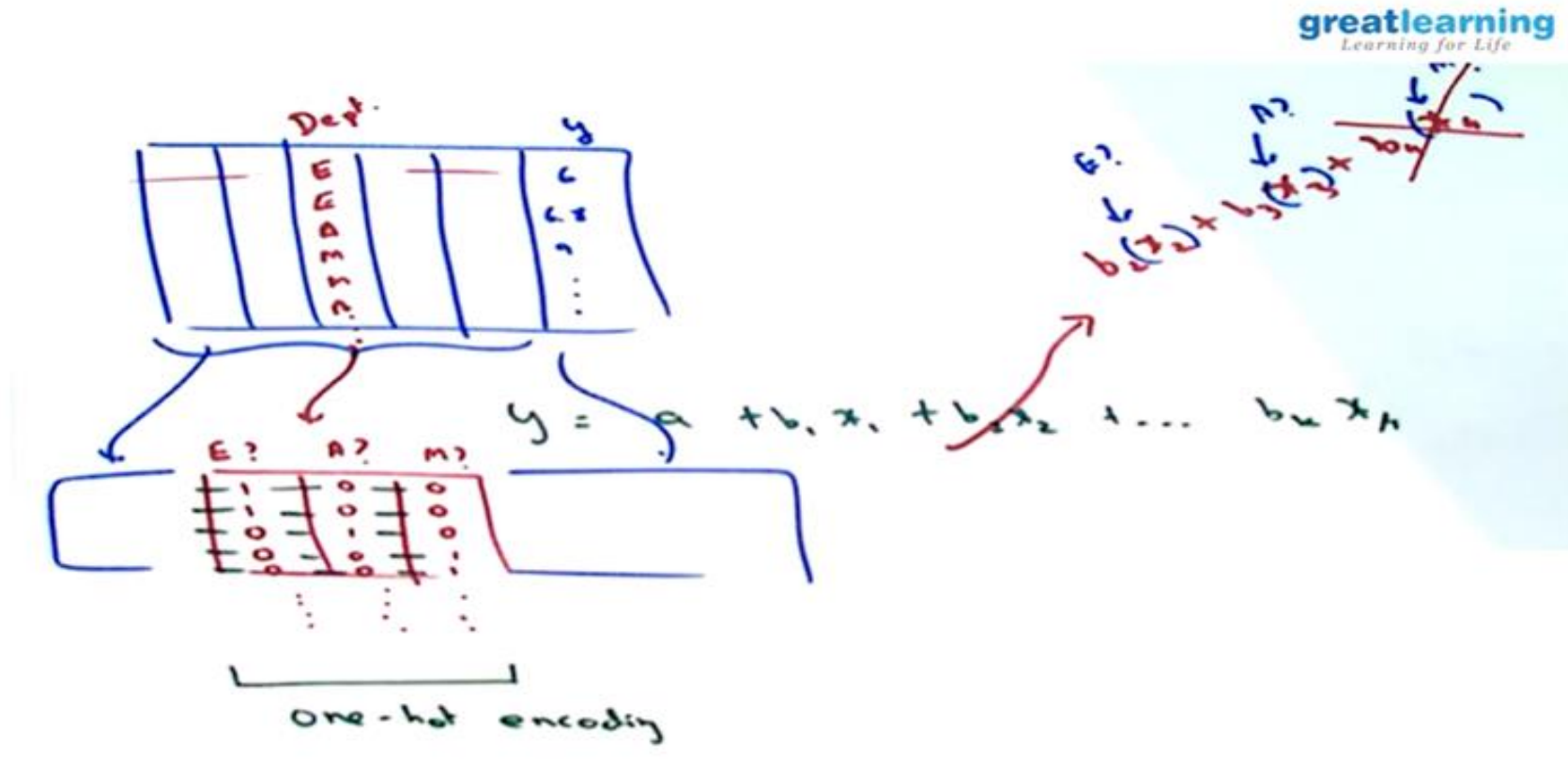
$R^2$  usually increase when var are added

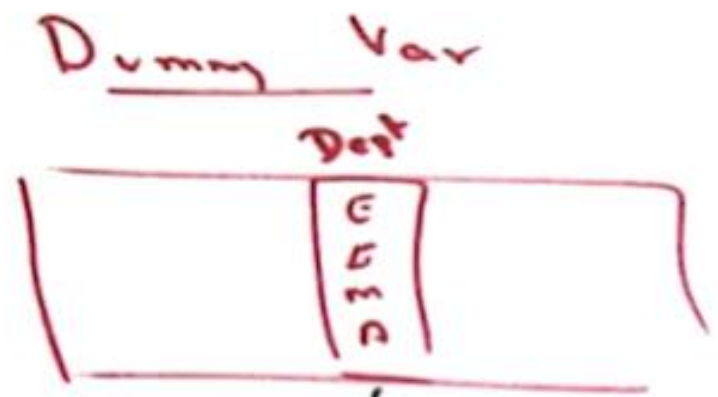
Adj  $R^2 \rightarrow$  will go up if ~~add~~ var is adding value

but will go down if var is not adding enough value

# One – Hot - Encoding

Categorical Independent variables in Regression



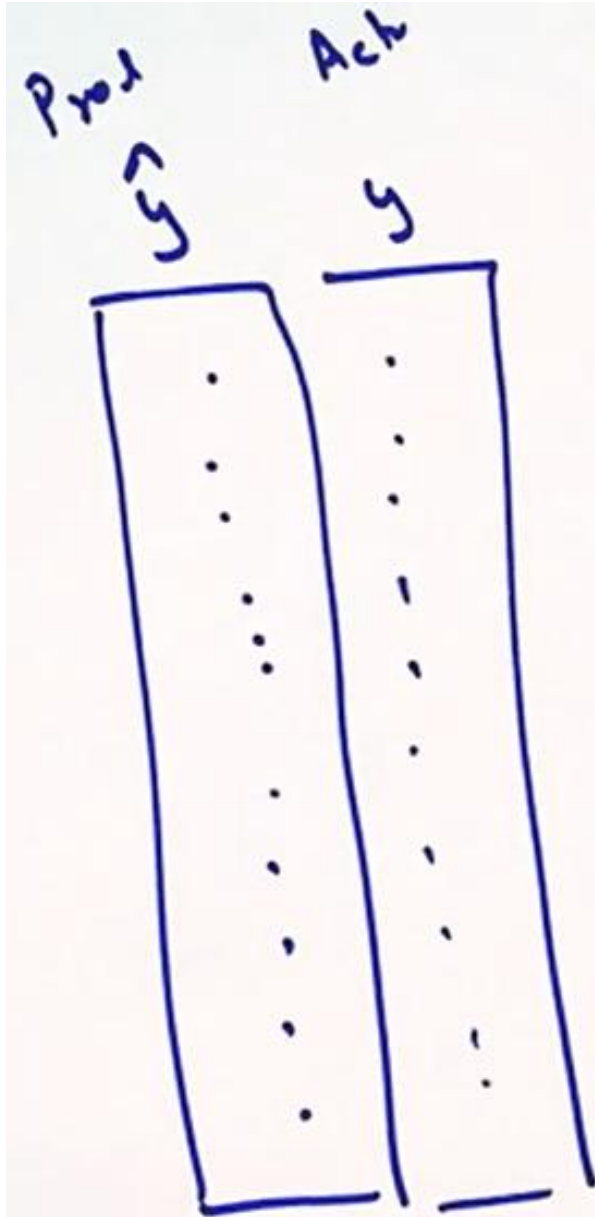


→ a categorical var with n possibilities → create n dummy var

$$y = a + \dots + b_3 \overbrace{(\text{Dept} = E)}^{x_3} + \underbrace{b_4}_{x_4} (\text{Dept} = A) + \dots$$

$b_4$  represents the diff from base (Dept = m)

Include only n-1 of these dummies in our reg.  
(the other is base)



$$\frac{1}{2} \sum |y_i - \hat{y}_i|$$

**MAE** **greatlearning**  
Learning for Life

**RMSE**

$$\sqrt{\frac{1}{N} \sum (y_i - \hat{y}_i)^2}$$



# Performance on Testing Data

- Important to keep in mind that all these performance measures will improve on training data as the model becomes more complex
- So these performance measures are best compared on testing data.
- Cross Validation is a commonly used to reduce over-fit and improve performance on test data



## Pros and Cons

- Advantages
- Simple elegant model
- Computationally very efficient
- Easy to interpret the output's coefficients

- Disadvantages

- Sometimes its just too simple to capture real-world complexities
- Assumes a linear relationships between dependent and independent variables.
- Outliers can have a large effect on the output ←
- Assumes independence between attributes

Handwritten notes illustrating a linear regression model:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3$$

Example calculation:

$$mpg = 46.317 - (0.0077)(wt)$$

Where  $b_1 = -0.0077$  is identified as  $b_1$ .

A small graph shows a downward-sloping curve, indicating a non-linear relationship.