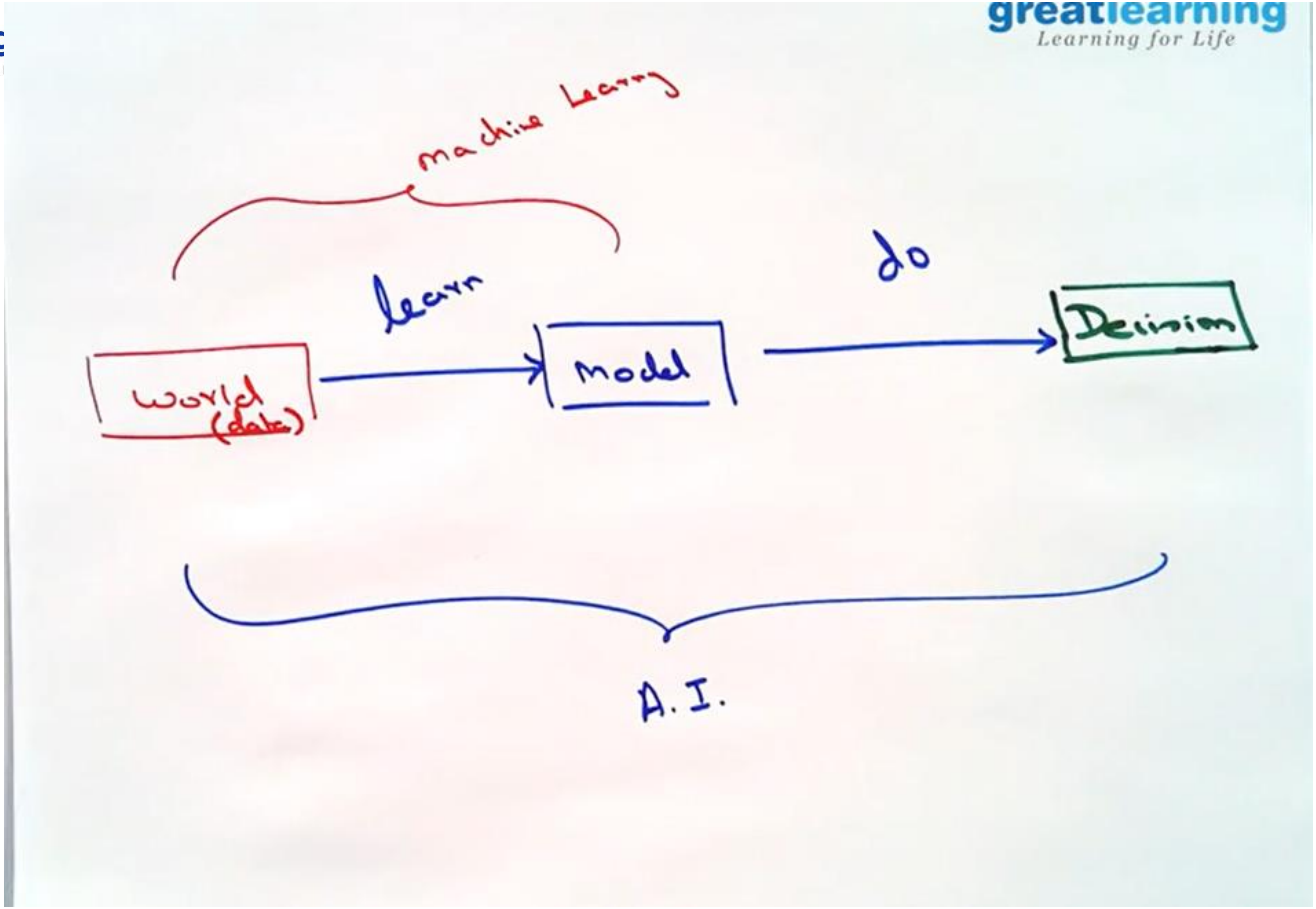
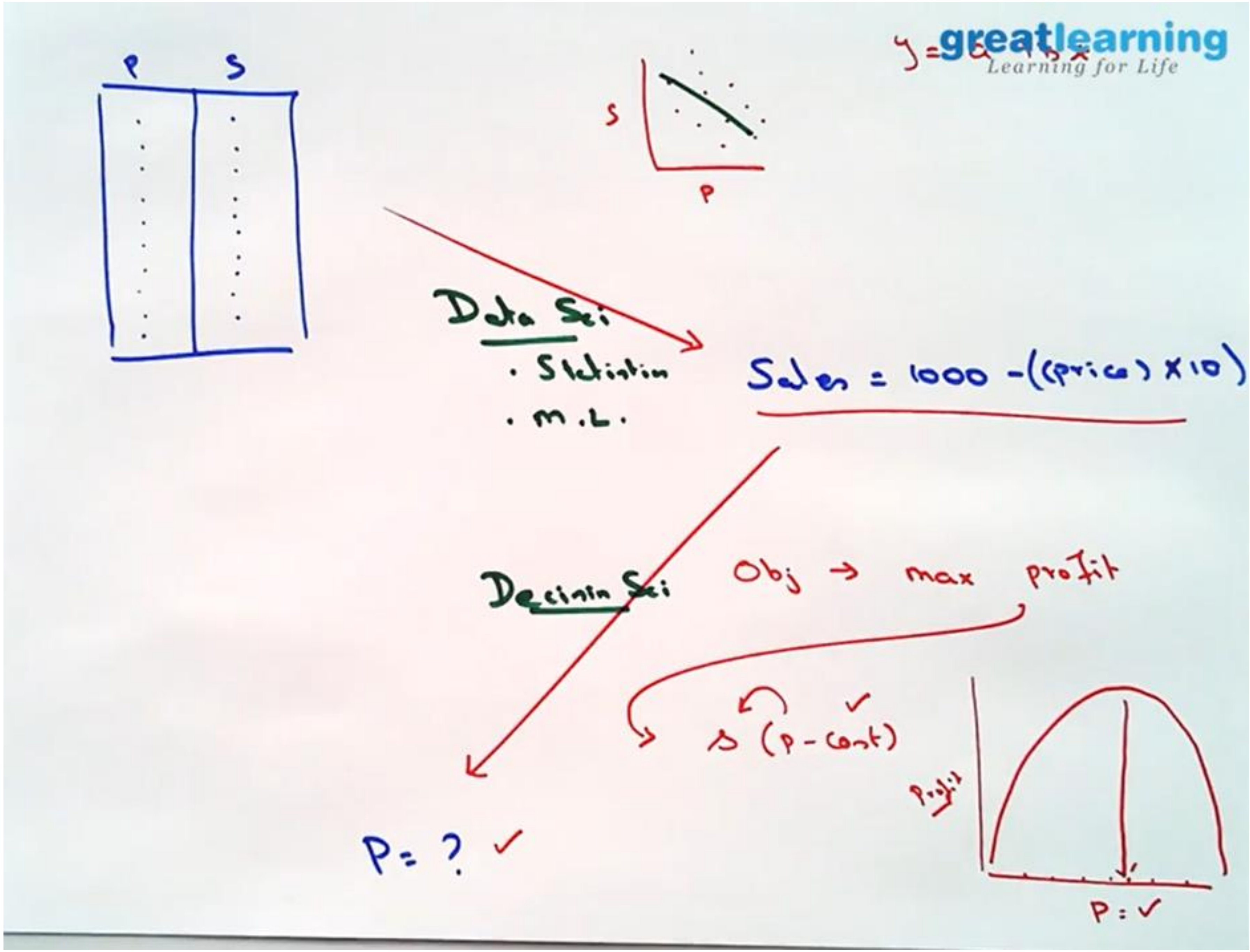
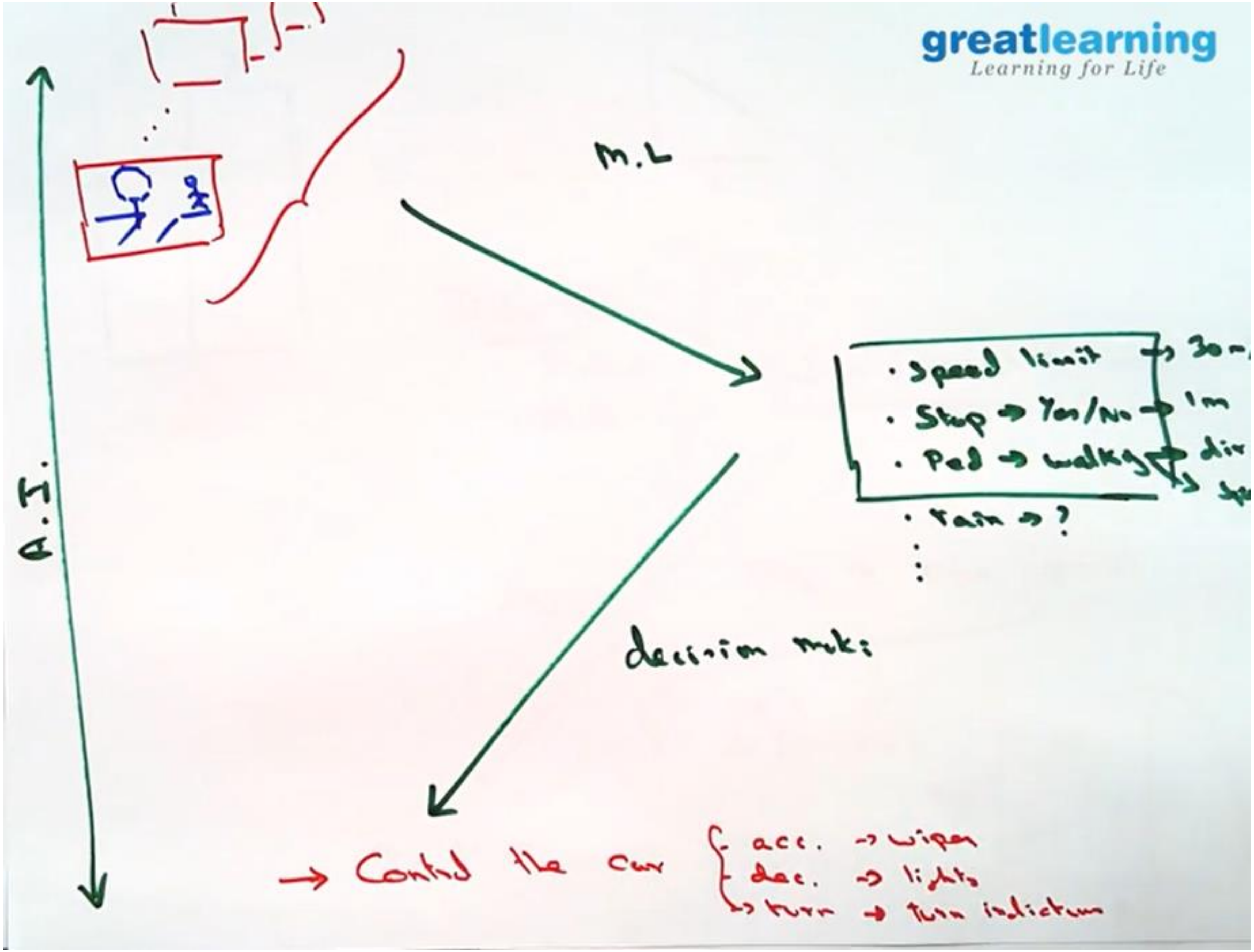
Intro to Artificial Intelligence and Machine Learning

does something Smart.


greatlearning
Learning for Life
machine learning
learn
do
World
(data)
model
Decision
A.I.


P
S
S
P
greatlearning
Learning for Life
Data Sci
· Statistics
· M.L.
Sales = 1000 −((price) x 10)
Decision Sci
Obj → max profit
S (P − Cost)
Profit
P = ?
P : ✓

Great Learning
POWER AHEAD



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

Great Learning
POWER AHEAD

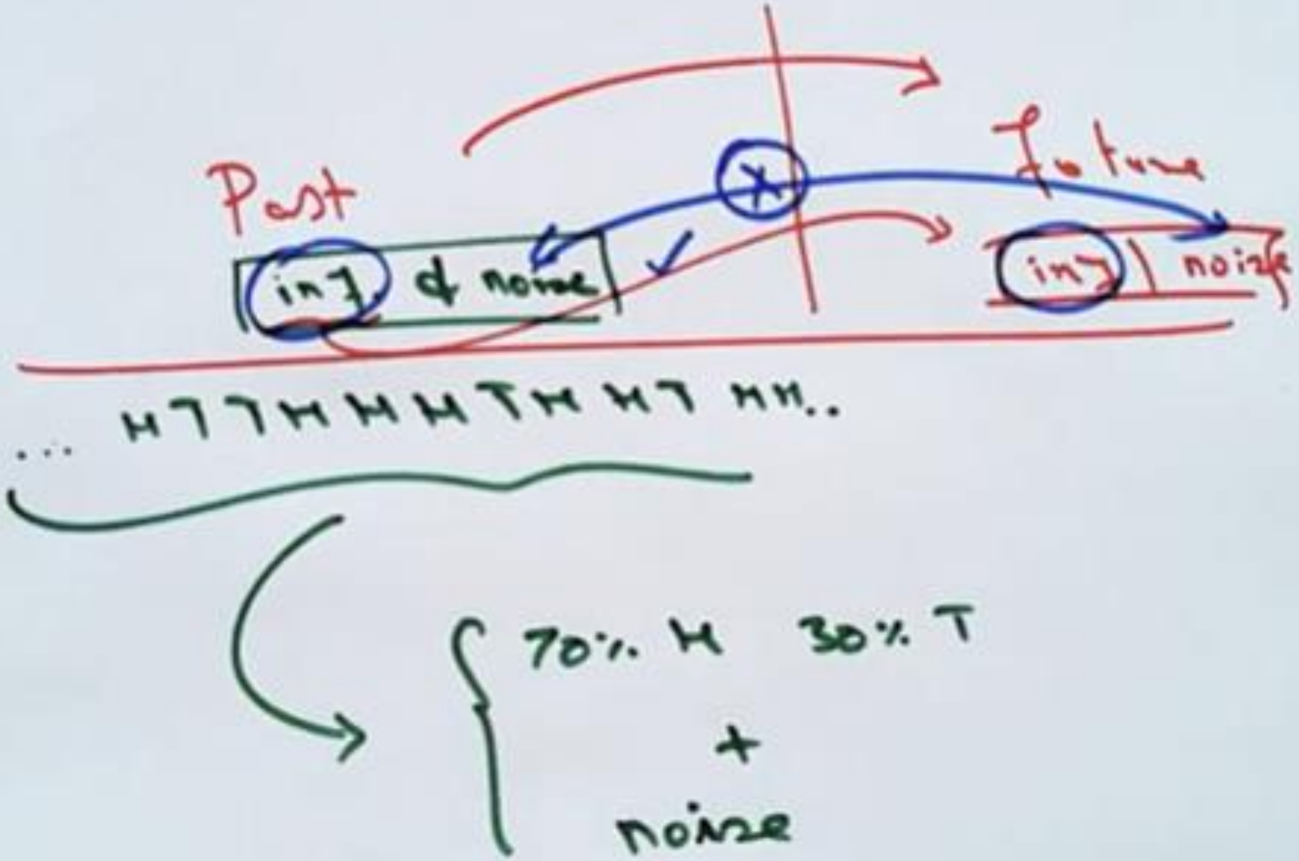
greatlearning
Learning for Life

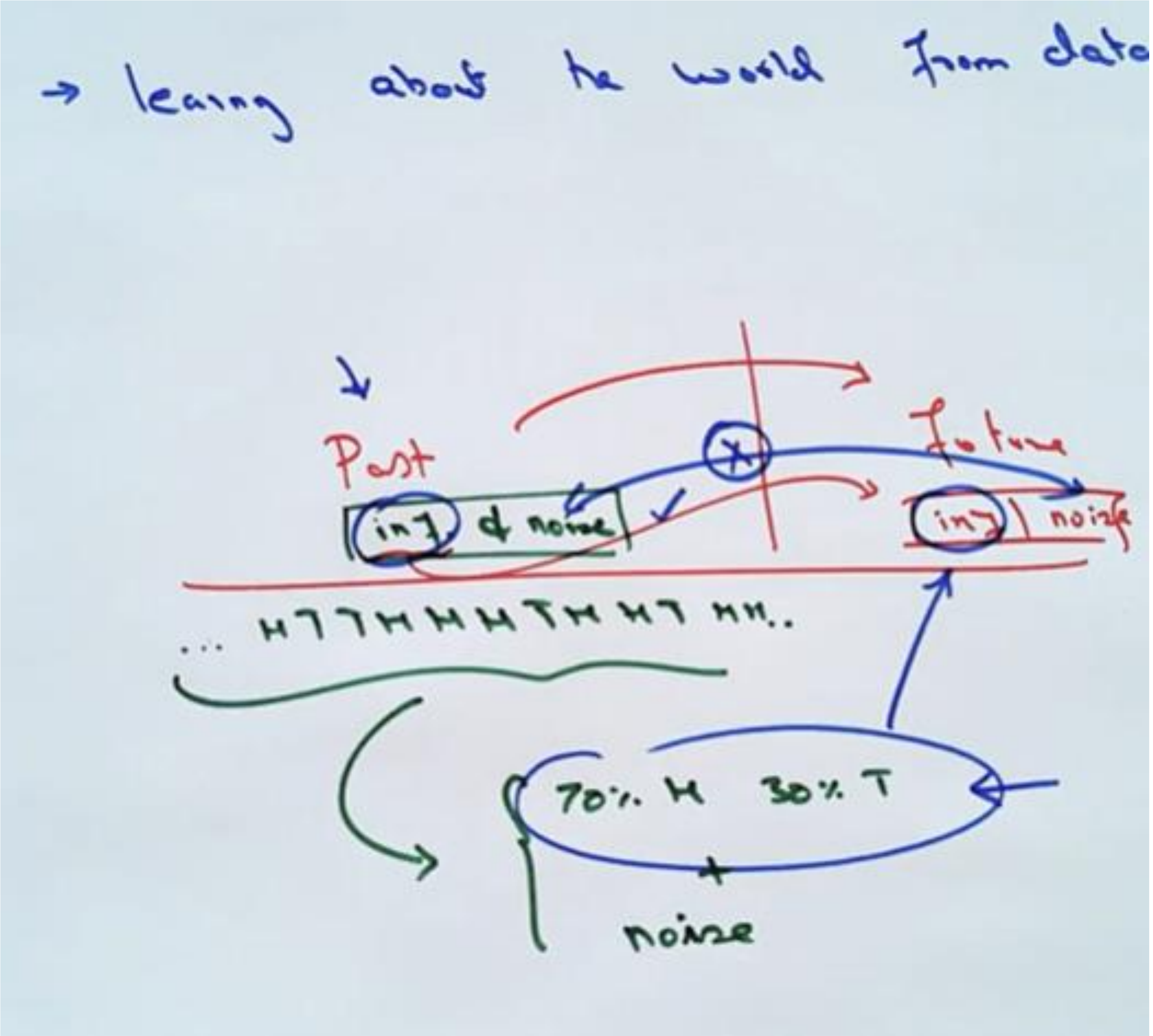
Program Overview: Machine Learning

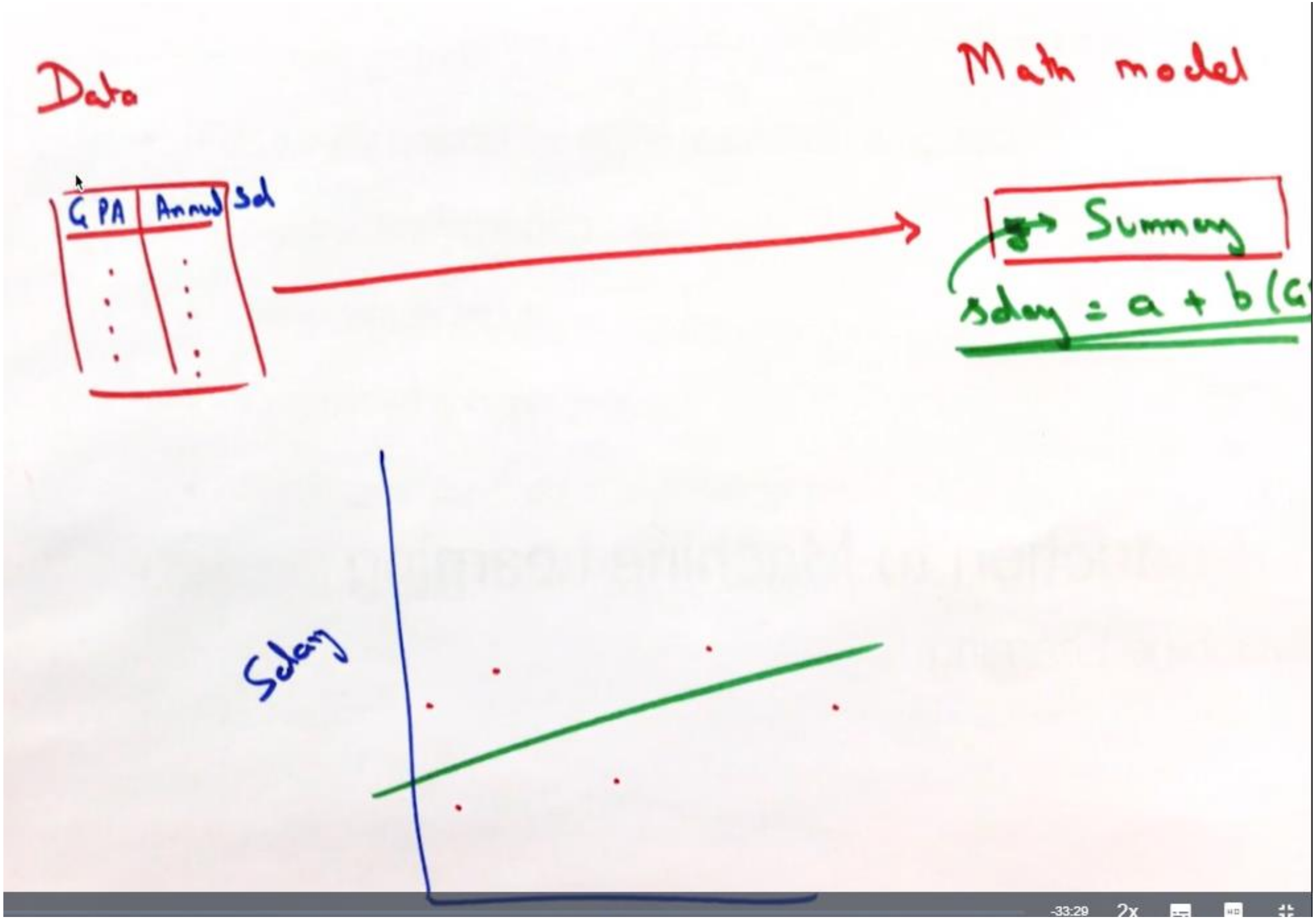
- Supervised Learning
- Unsupervised Learning
- Ensemble Techniques
- Recommendation Systems

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

→ learning about the world from data



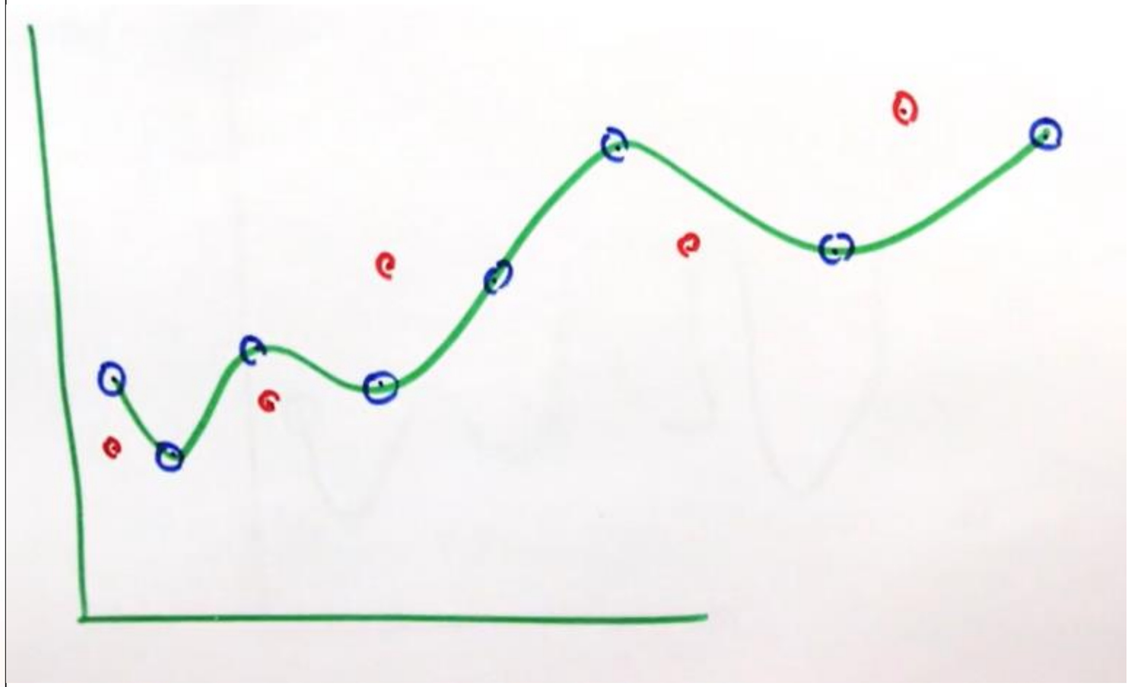

→ learning about the world from data
Past
Future
inf & noise
inf | noise
... H T T H H H H T H H T H H..
70% H 30% T
+
noise

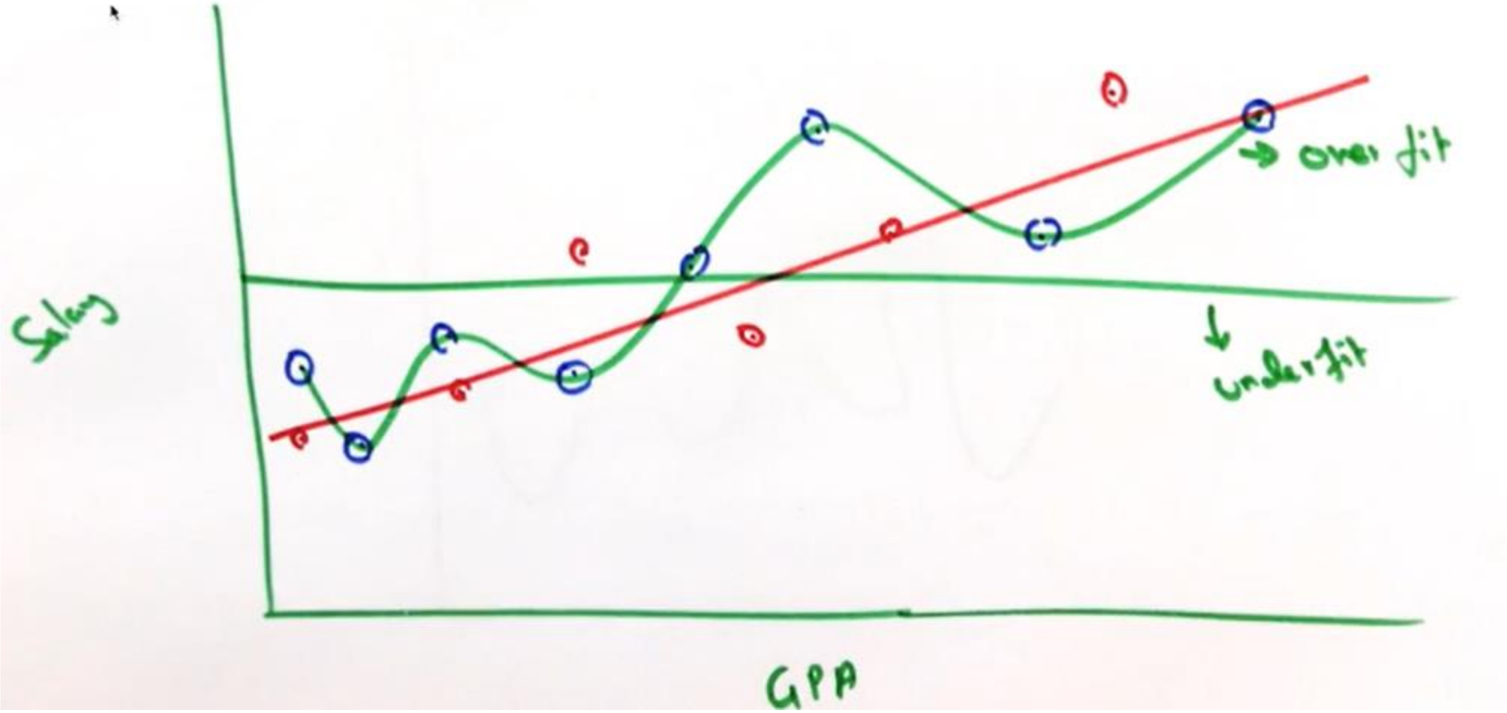

Data
Math model
GPA
Annual Sal
Summary
Salary = a + b (G
Salary
-33:29
2x

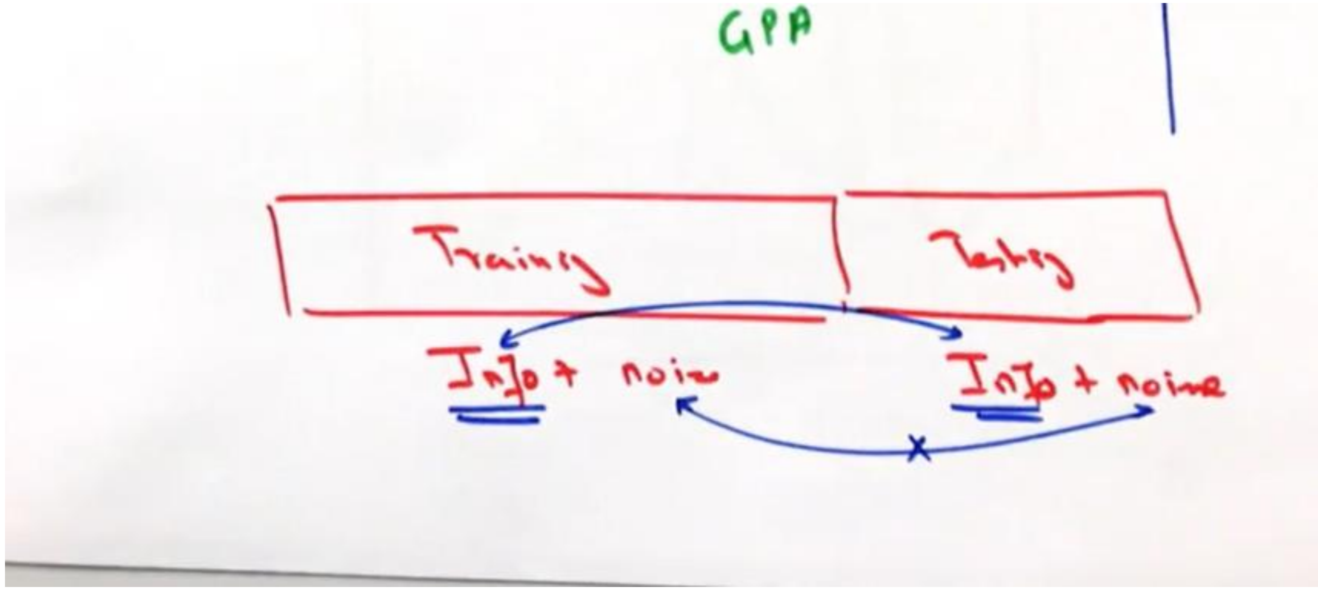
Learning from Data

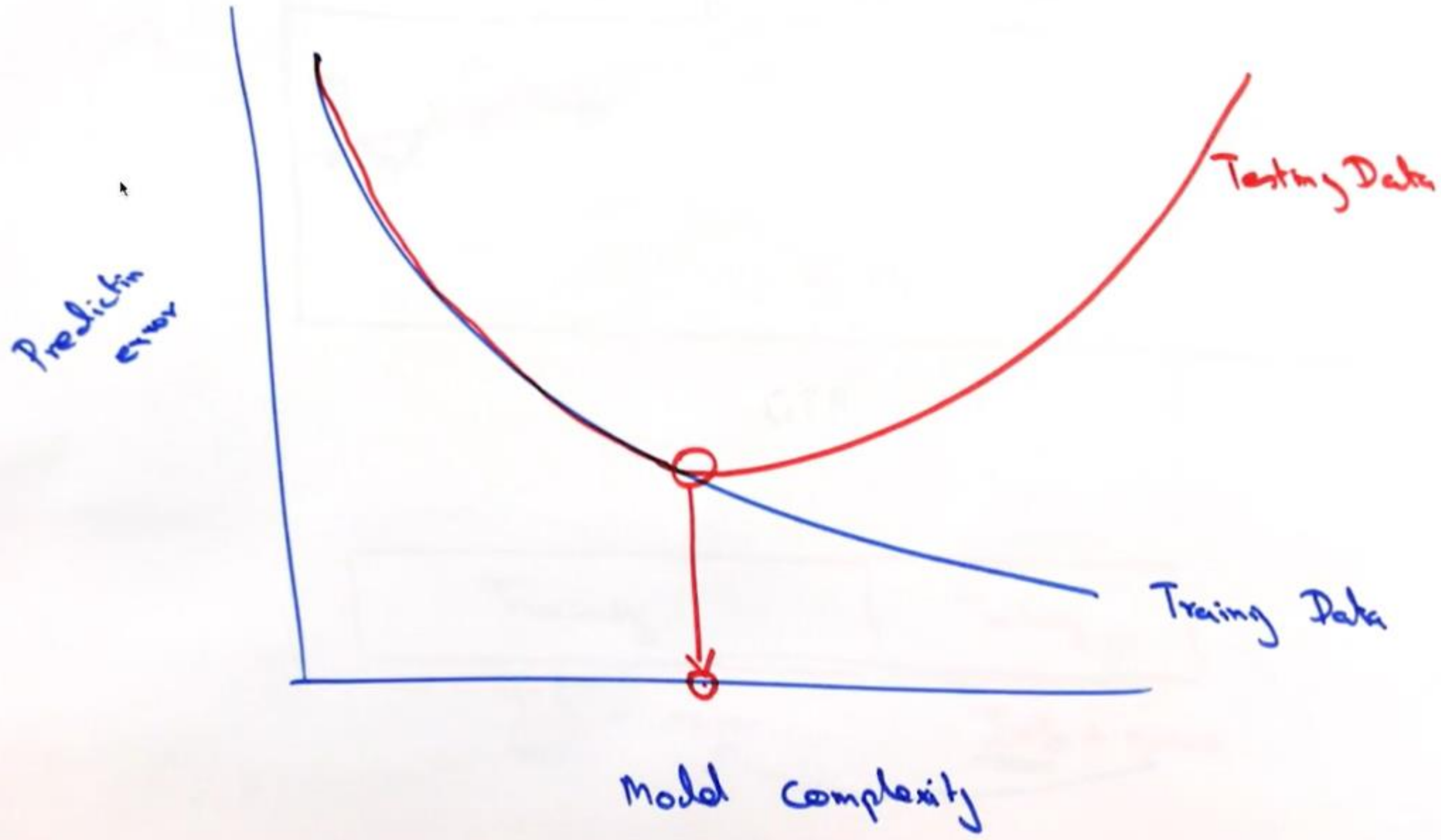
- Can we learn about the world around us using data?
- Model building from data
  - Take data as input
  - Find patterns in the data
  - Summarize the pattern in a mathematically precise way
- Machine learning automates this model building.

Great Learning
POWER AHEAD




over fit
Salary
underfit
GPA


GPA
Training
Testing
Info + noise
Info + noise


Testing Data
Prediction error
Training Data
Model complexity

The Challenge

- Data unfortunately contains noise. If not, machine learning would be trivial!
- Think of Data = Information + 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.

Great Learning
POWER AHEAD

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.

-15:55 2x

Supervised Learning

Unsupervised Learning

Great Learning
POWER AHEAD

greatlearning
Learning for Life

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 is 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.

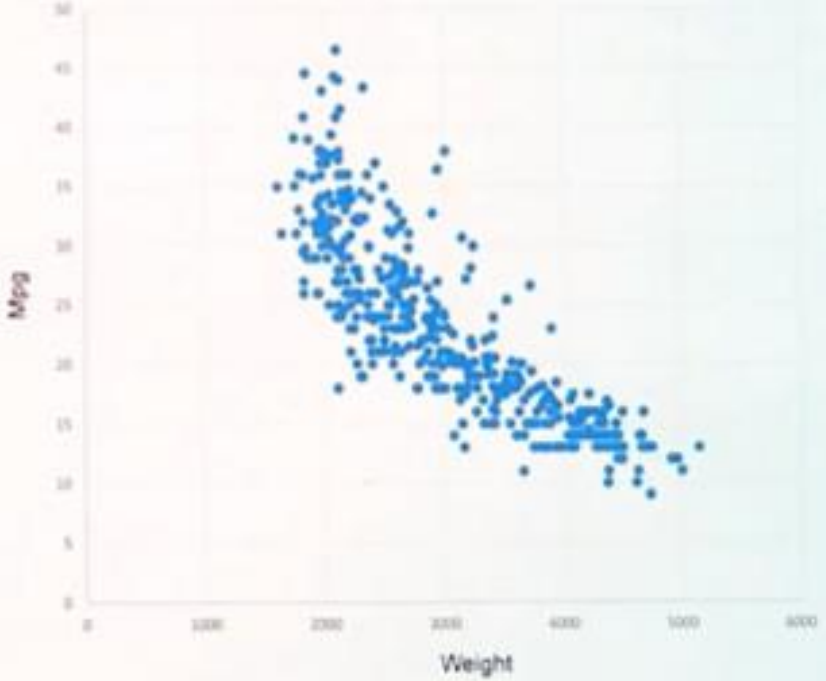
- Supervised learning
  - Regression: desired output is a continuous number
  - Classification: desired output is a category
- Unsupervised learning
  - Clustering: Grouping data
  - Dimensionality reduction: Compressing data
  - Association rule learning: If X then Y

Linear Relations between two variables

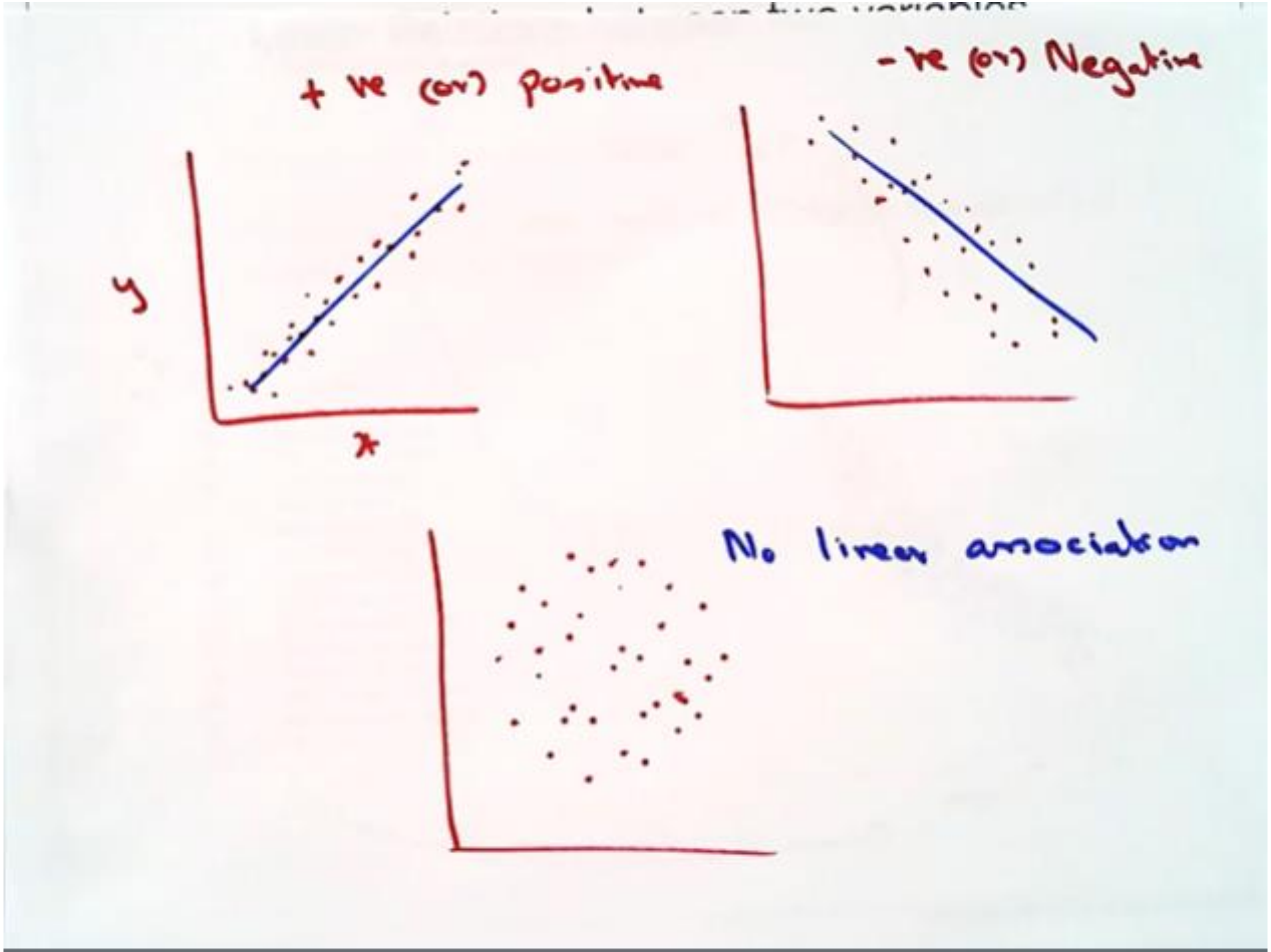
- Do heavier cars have lower mileage?
- Can we use DATA to better understand relationships between the two variables: weight and mpg?

y x

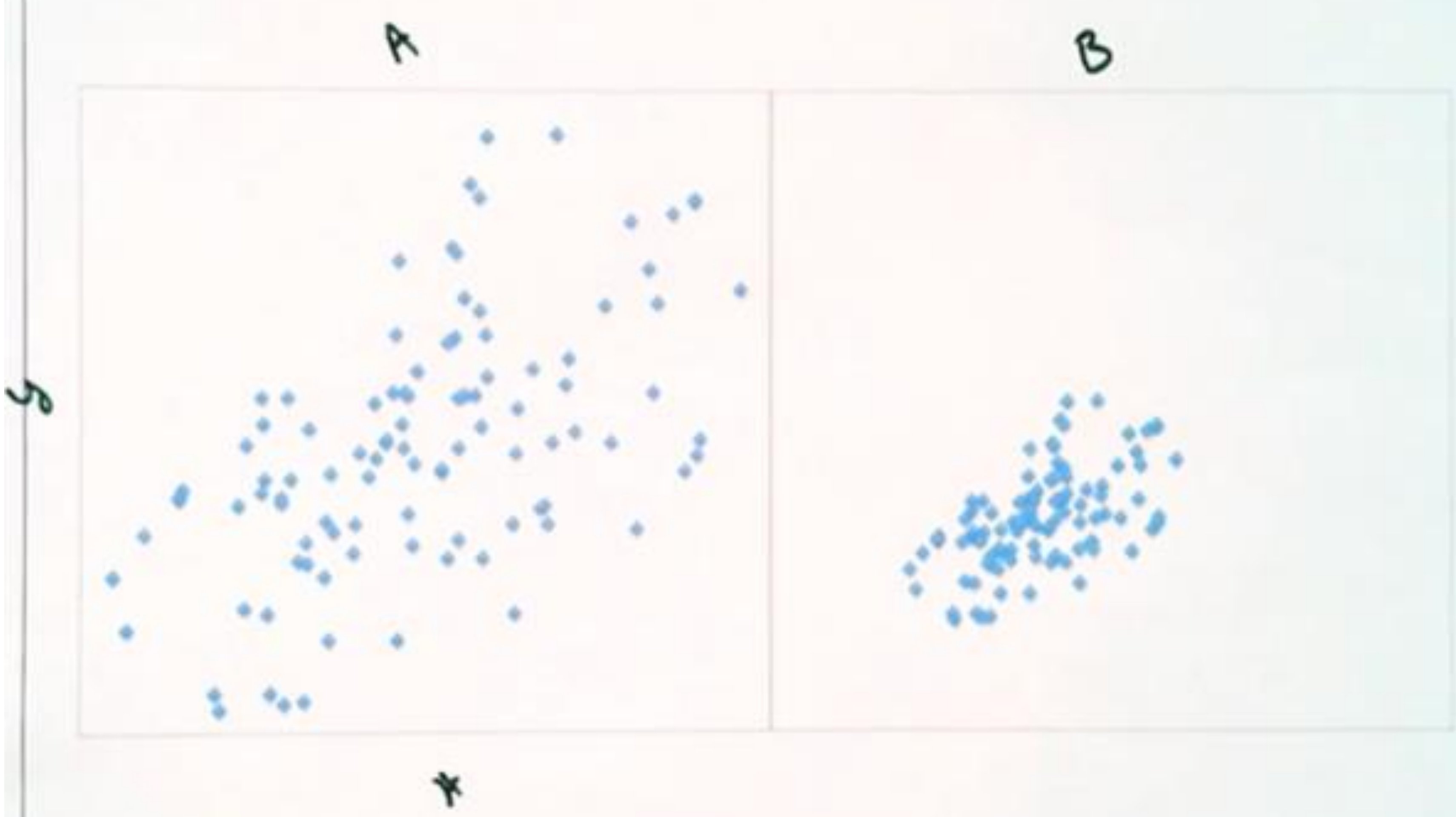
mpg	wt	car_name
18	3504	chevrolet chevelle malibu
15	3693	buick skylark 320
18	3436	plymouth satellite
16	3433	amc rebel sst
17	3449	ford torino
15	4341	ford galaxie 500
14	4354	chevrolet impala
14	4312	plymouth fury iii
14	4425	pontiac catalina
15	3850	amc ambassador dpl
15	3563	dodge challenger se
14	3609	plymouth 'cuda 340
15	3761	chevrolet monte carlo
14	3086	buick estate wagon (sw)
24	2372	toyota corona mark ii

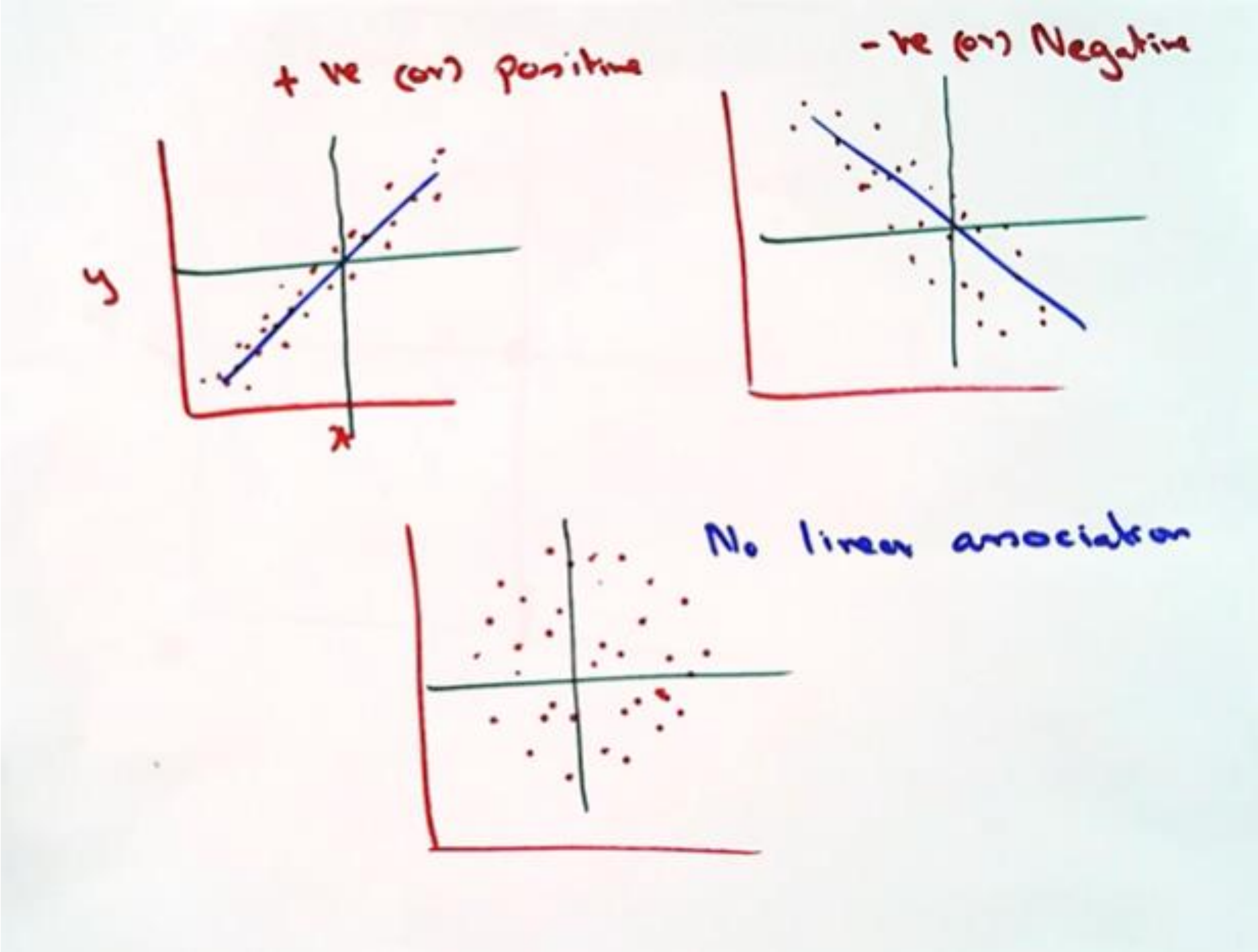


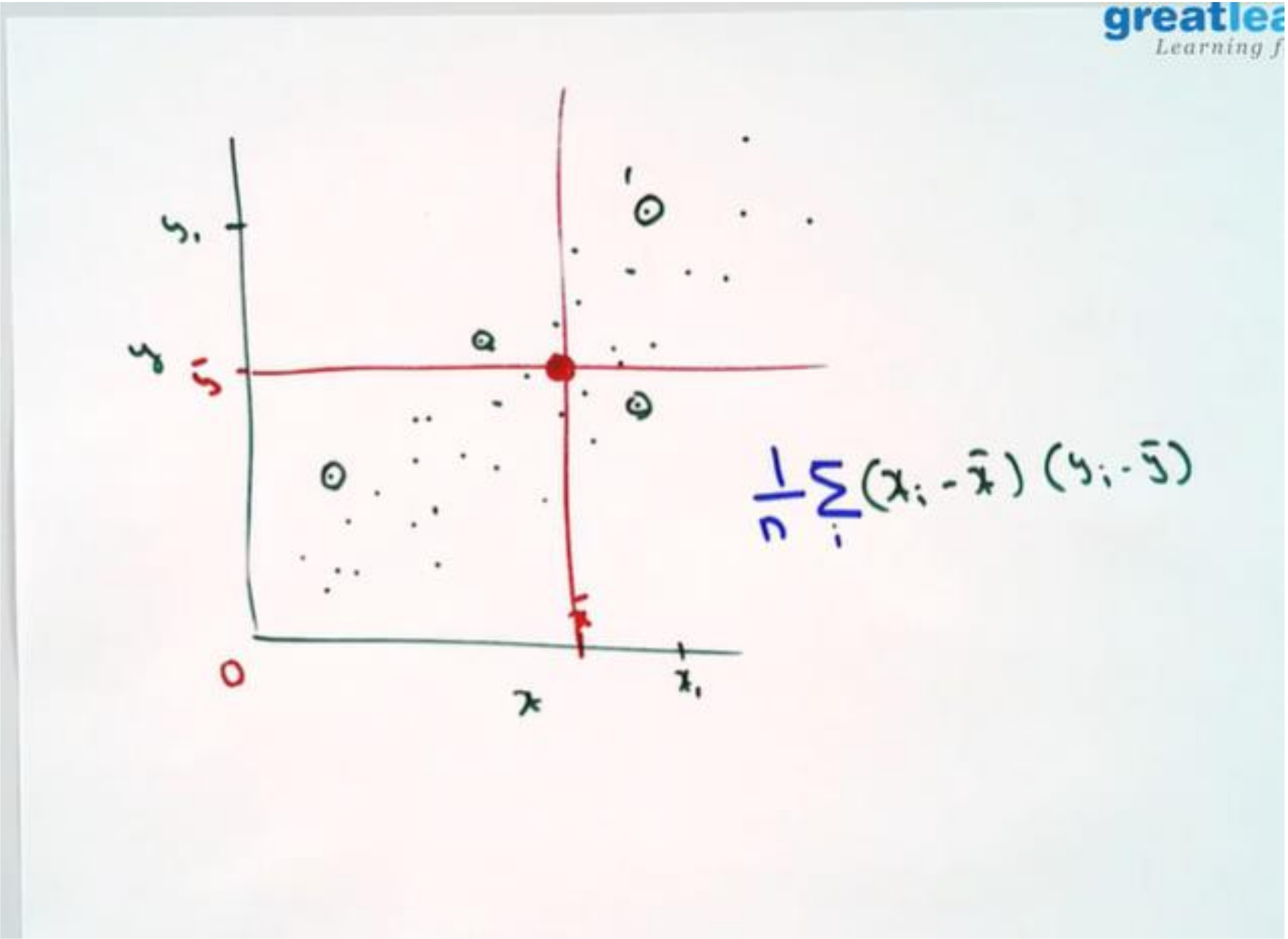
Data Source: StatLib (http://lib.stat.cmu.edu/datasets/)


+ve (or) positive
y
x
-ve (or) Negative
No linear association

Which one has a stronger relationship?




+ve (or) positive
y
x
-ve (or) Negative
No linear association


y_1
y
$\bar{y}$
0
x
$\bar{x}$
x_1
$\frac{1}{n}\sum_i (x_i - \bar{x})(y_i - \bar{y})$

Great Learning
POWER AHEAD

greatlearning
Learning for Life

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 form 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 lets define "covariance" between X and Y as

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

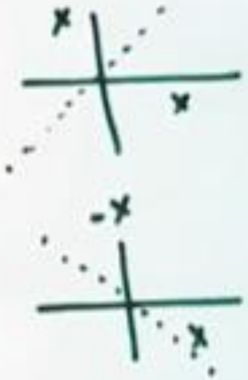
$$\cancel{(\text{unit } x)}\ \cancel{(\text{unit } y)}$$

$$Corr(x, y) = \boxed{\frac{Cov(x, y)}{Std\ dev(x)\ \ Std\ dev(y)}} \Rightarrow \text{doesn't have any unit}$$

$$\cancel{(\text{unit } x)} \qquad \cancel{(\text{unit } y)}$$

$$Corr(x, x) = 1$$

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



- 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

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

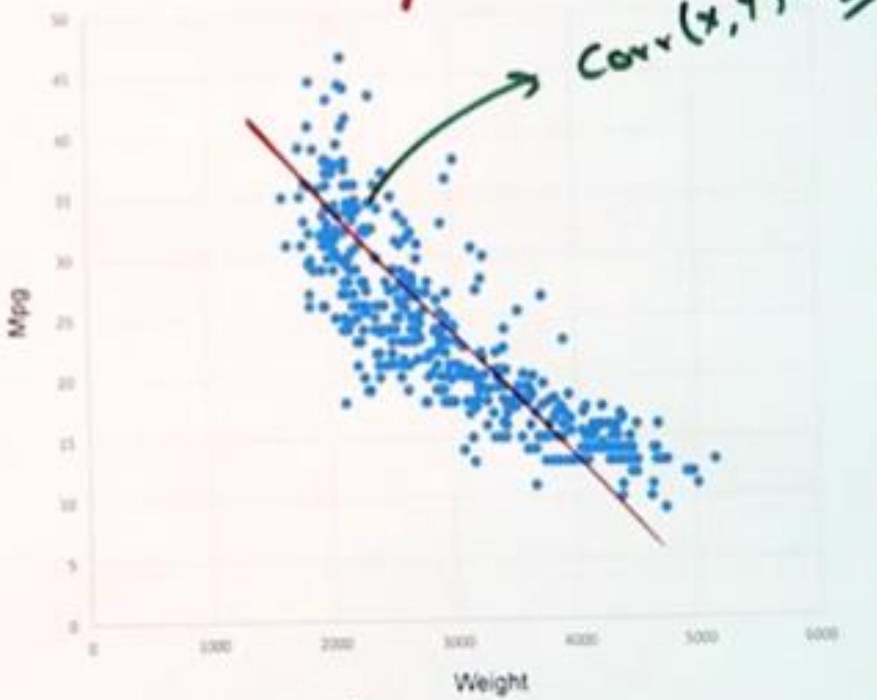
  - Correlation is always between -1 and +1.The correlation between a variable and itself is 1.
  - The correlation between X and Y is the same as the correlation between Y and X.
  - Correlation is scale invariant

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?

y x

mpg	wt	car_name
18	3504	chevrolet chevelle malibu
15	3693	buick skylark 320
18	3436	plymouth satellite
16	3433	amc rebel sst
17	3449	ford torino
15	4341	ford galaxie 500
14	4354	chevrolet impala
14	4312	plymouth fury iii
14	4425	pontiac catalina
15	3850	amc ambassador dpl
15	3563	dodge challenger se
14	3609	plymouth 'cuda 340
15	3761	chevrolet monte carlo
14	3086	buick estate wagon (sw)
24	2372	toyota corona mark ii



$Corr(x, y) = -0.83$

Data Source: StatLib (http://lib.stat.cmu.edu/datasets/)

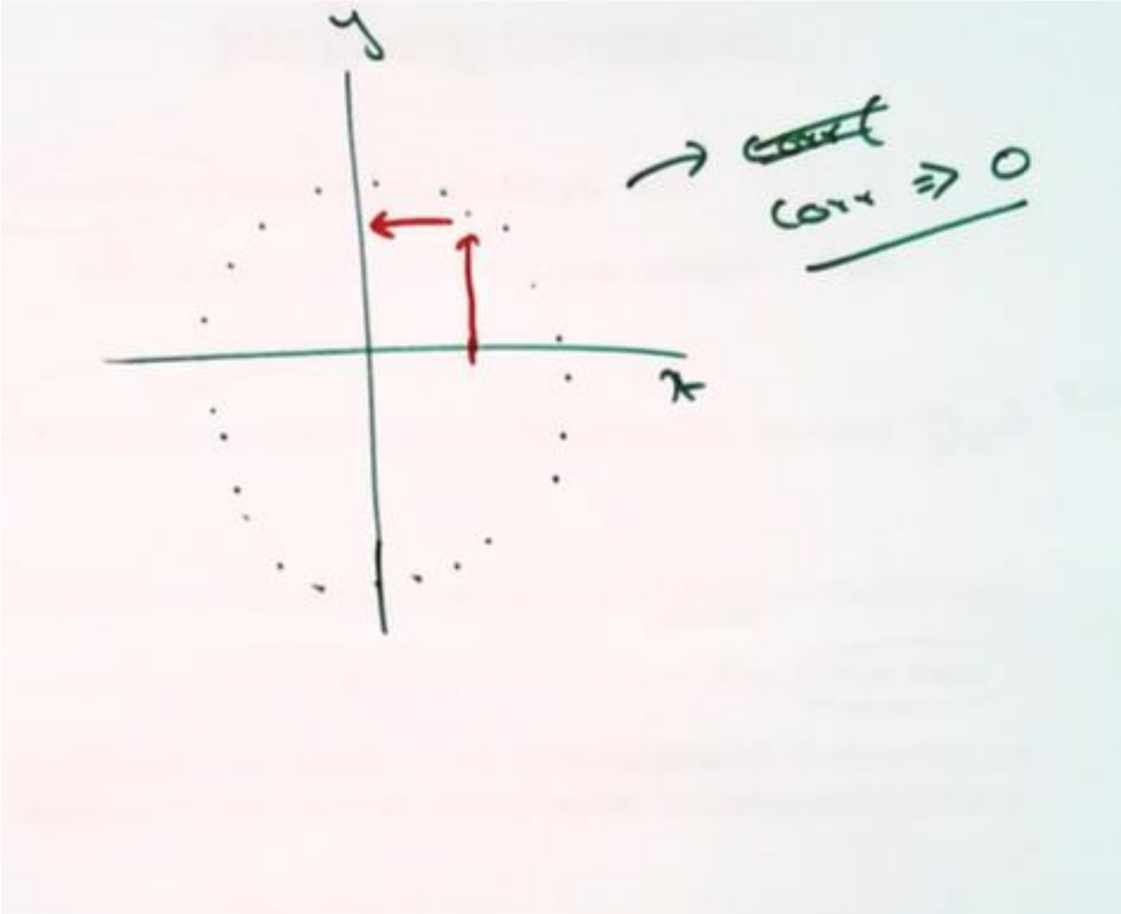
Interpreting Correlations

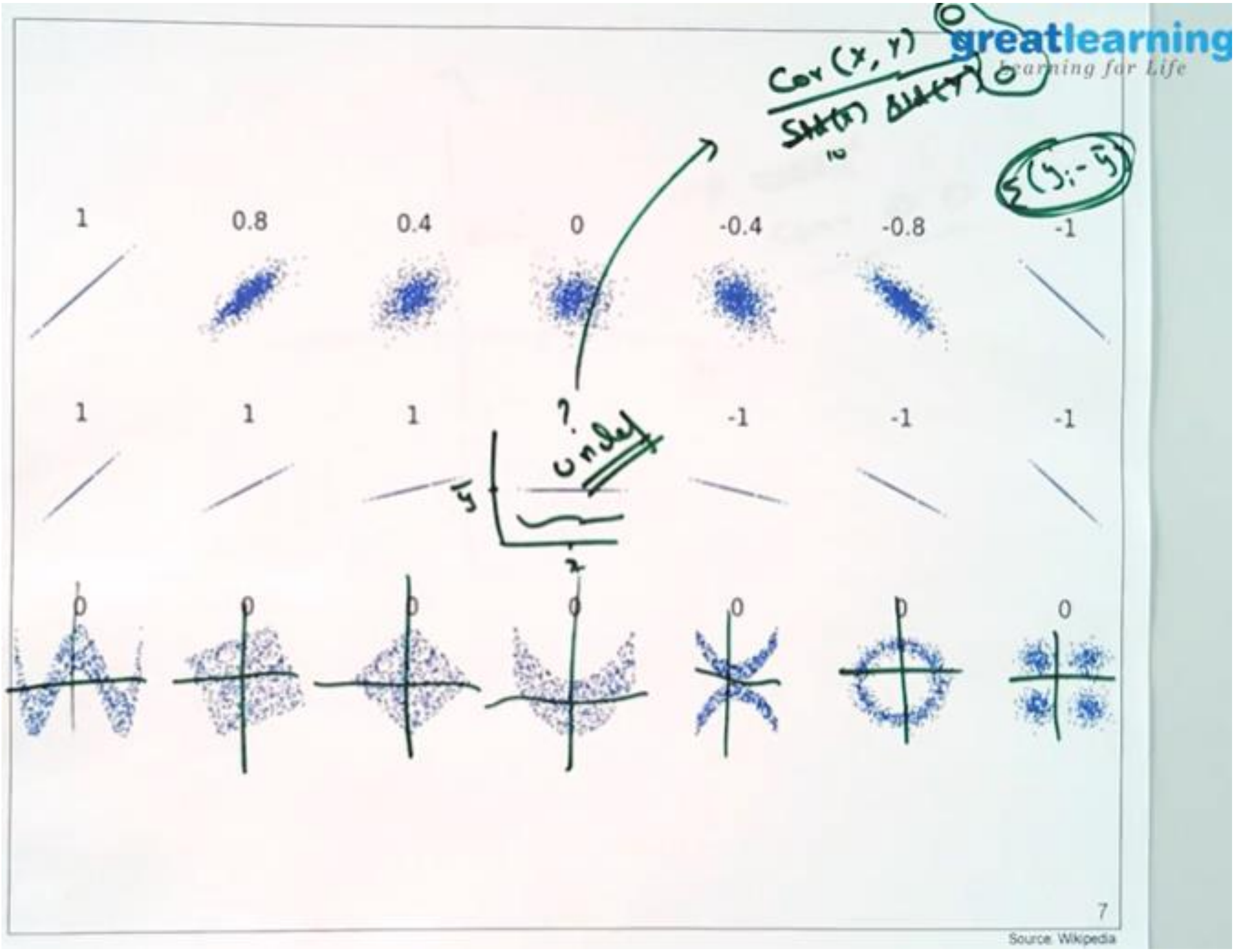
- Correlation between Weight and Mpg is -0.83
  - Does heavier car tend to have a lower mileage? Yes

  Cavsation

  - If we increase the weight of a car, will its Mpg decrease? Dont 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.

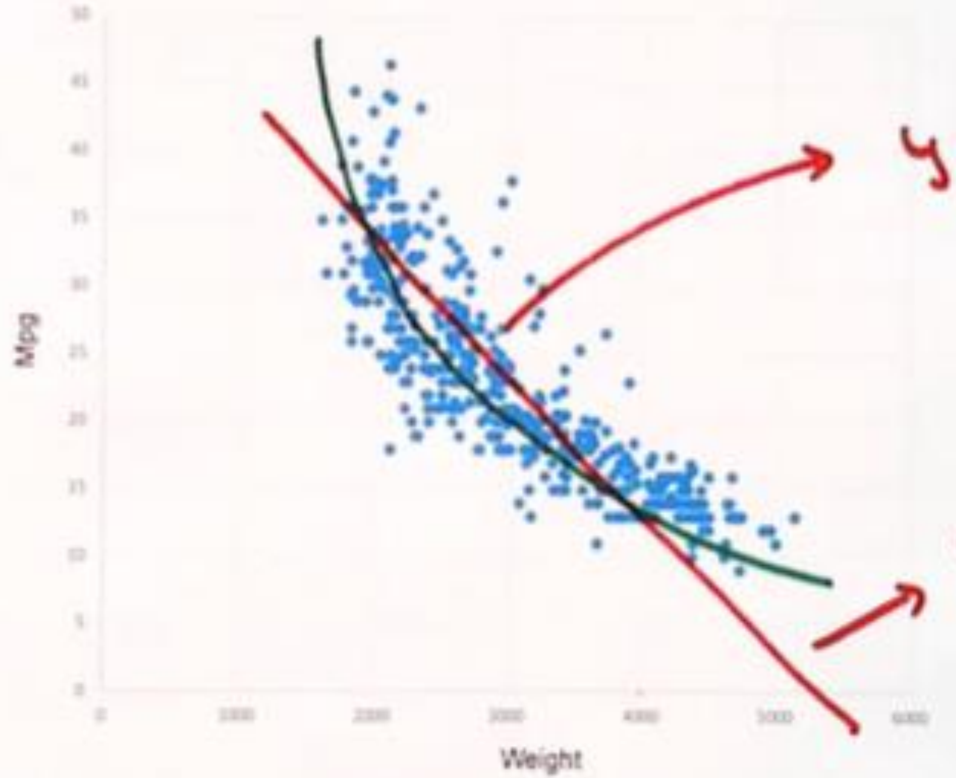

y
x
Corr ⇒ 0


Cov (x, y)
greatlearning
Learning for Life
1
0.8
0.4
0
-0.4
-0.8
-1
1
1
1
-1
-1
-1
0
0
0
0
0
0
0
7
Source: Wikipedia

- Next: If a car's weight is 4000, what would we expect its Mpg to be?
  - Previously: Measuring strength of relationship

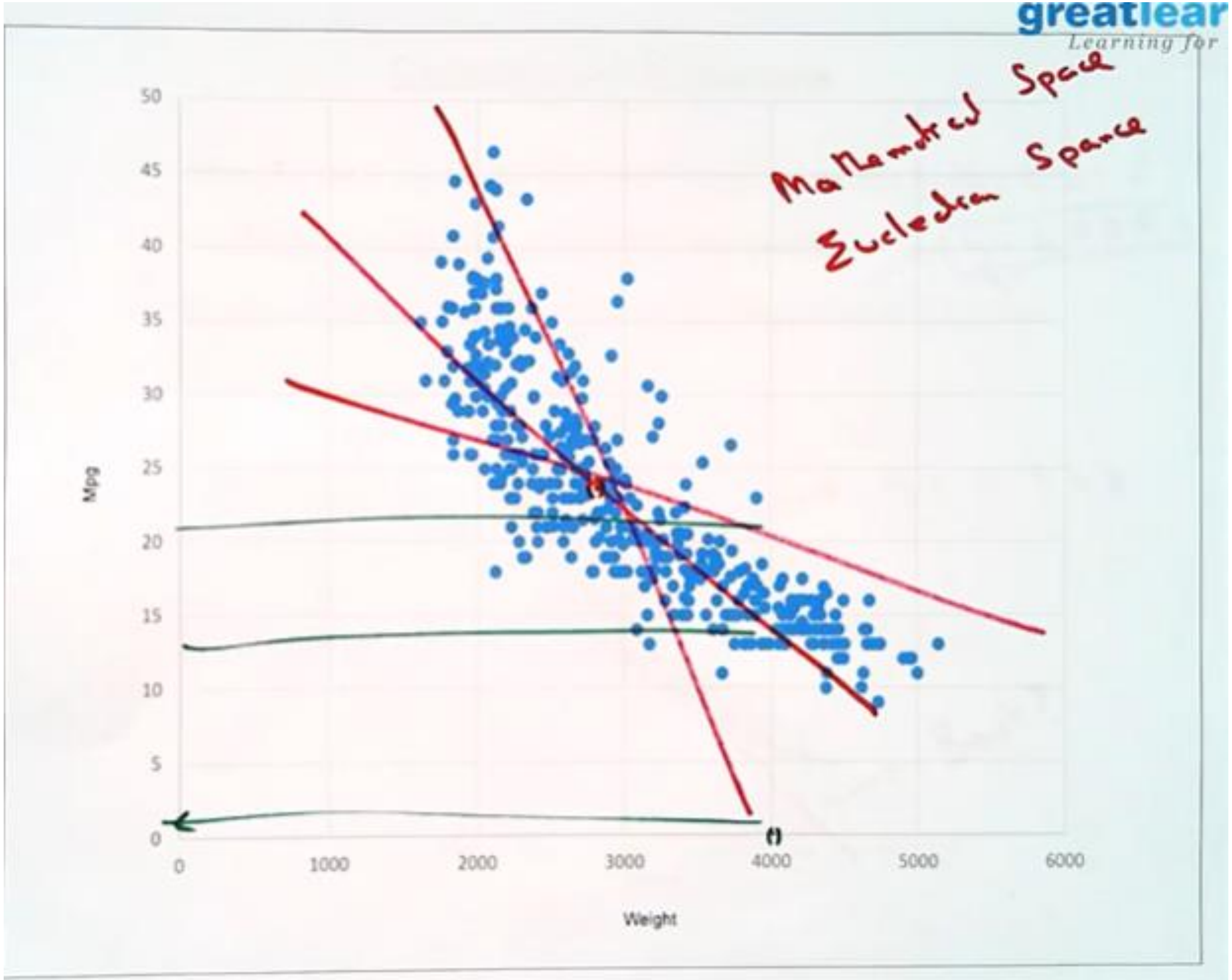
$\boxed{y = a + bx}$

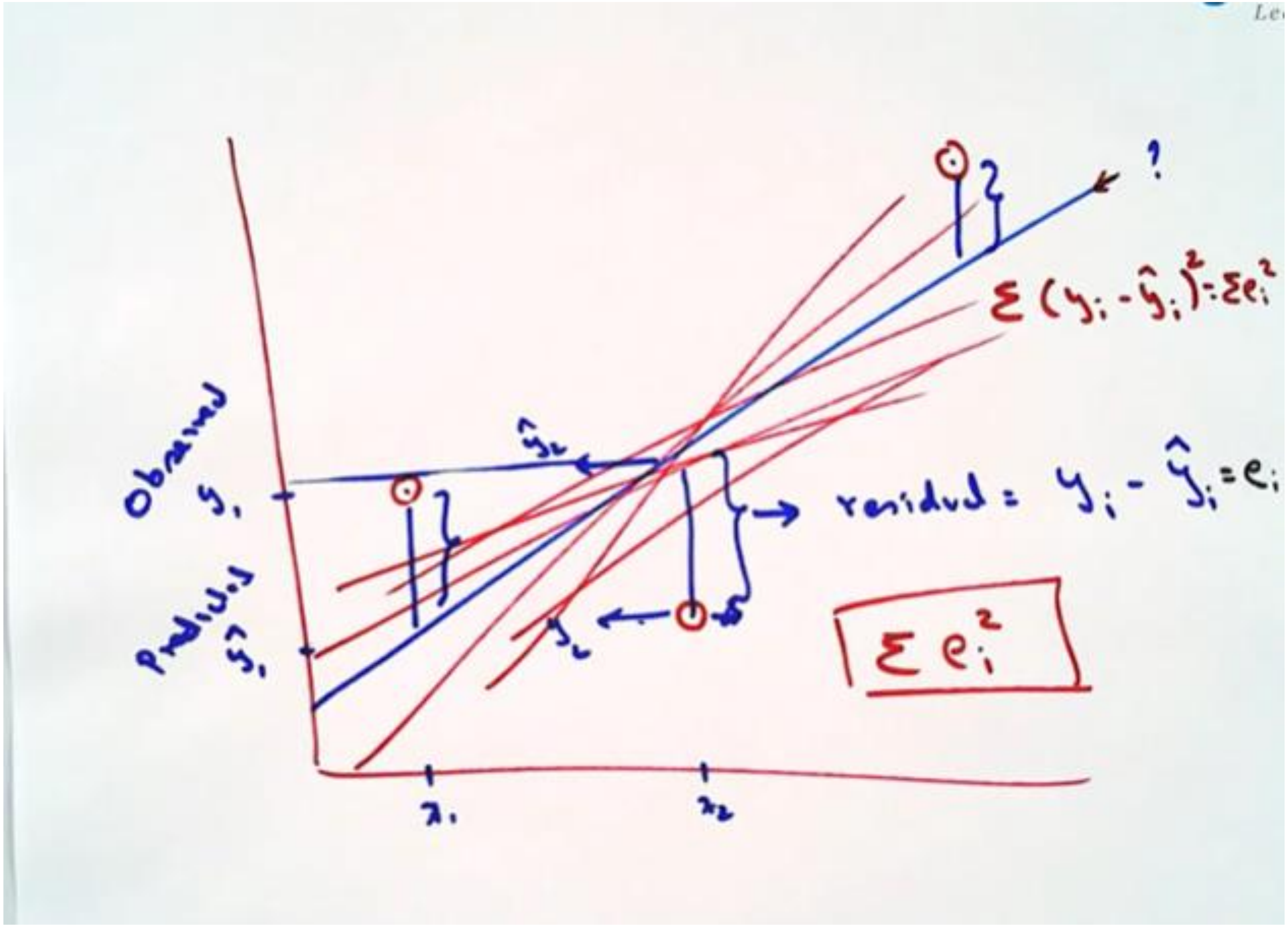
  - Now: Capturing relationships using a simple model (equation)



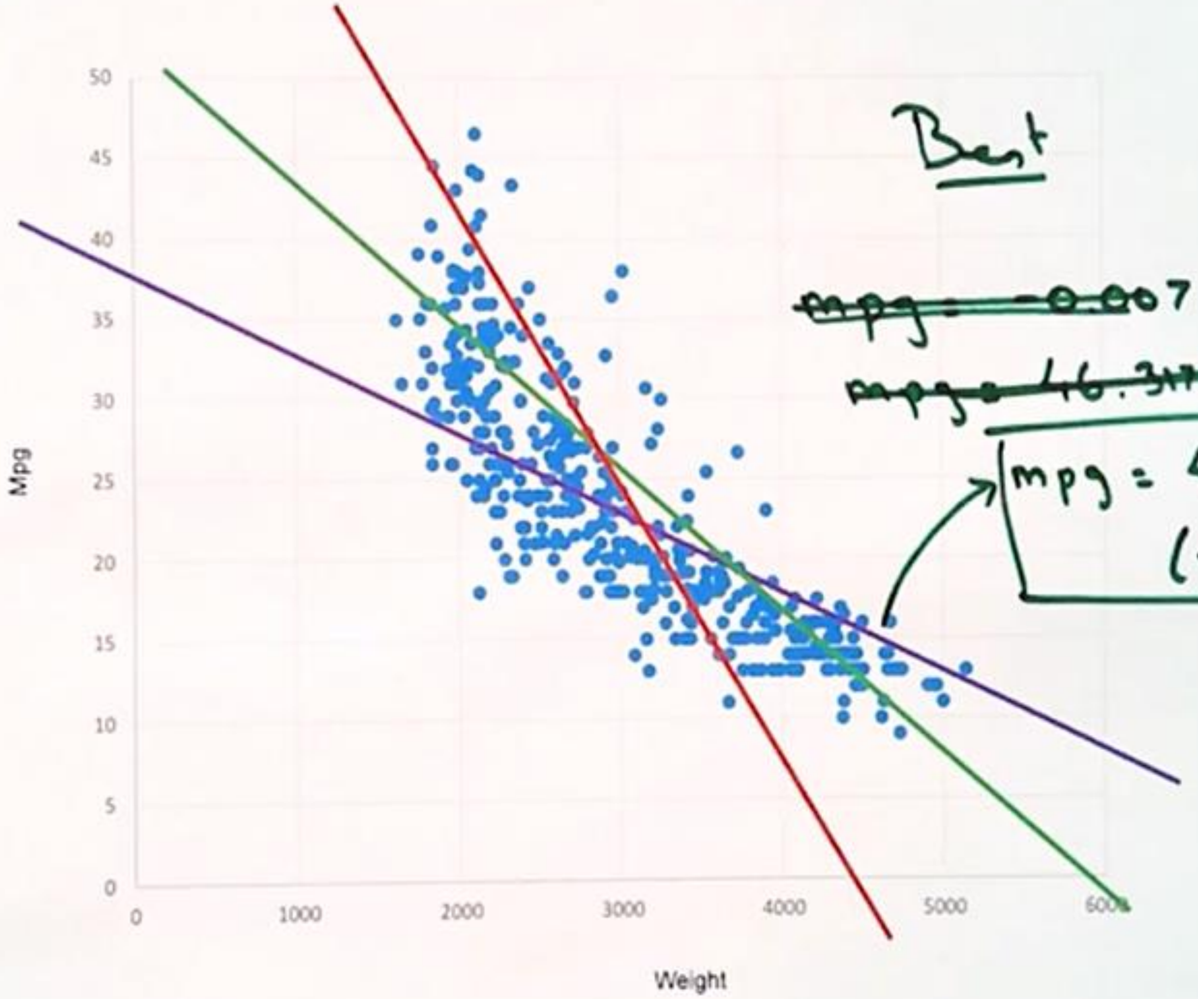
$y = a + bx$

Best?


Mathematical Space
Eucledian Space
Mpg
Weight
50
45
40
35
30
25
20
15
10
5
0
0
1000
2000
3000
4000
5000
6000


Observed y_1
Predicted $\hat{y}_1$
$\hat{y}_2$
y_2
residual = $y_i - \hat{y}_i = e_i$
$\Sigma(y_i - \hat{y}_i)^2 = \Sigma e_i^2$
?
Σe_i^2
x_1
x_2

How easy is it to fit a straight line?



Best

~~mpg = -0.007~~

~~mpg = 46.317 - 0.0077X~~

$$\boxed{mpg = 46.317 + (-0.0077\,wt)}$$

So...

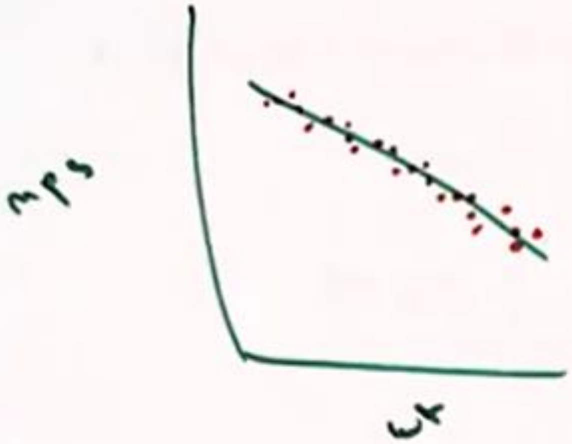
- If a car's weight is 4000, what would we expect its Mpg to be?

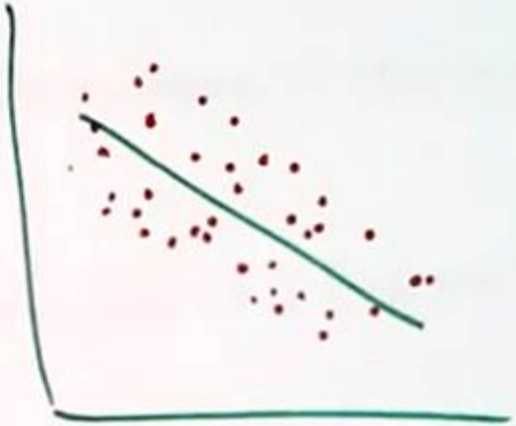
$$mpg = 46.317 - (0.0077)(4000)$$

$$= \boxed{15.52}\ mpg$$

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

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


mpg
wt



How good is our regression fit?

Std. dev of residuals = Std. error of the Regression line

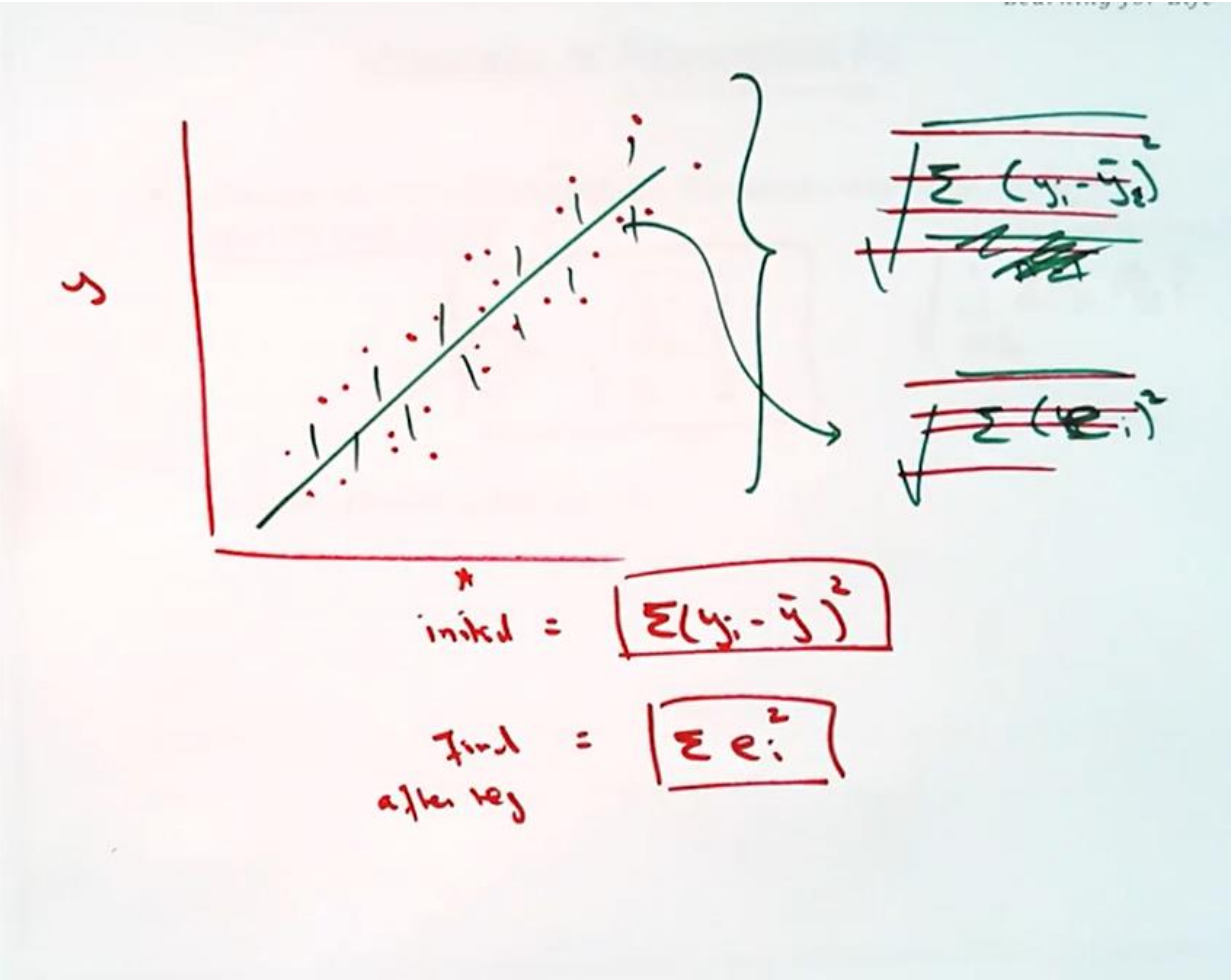
Measures of Regression Fit

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

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

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

- Comparing RMSE to Std. dev of y



y
initial = Σ(yᵢ - ȳ)²
final after reg = Σ eᵢ²

$$R^2 = 1 - \frac{\overbrace{\sum e_i^2}^{SSE}}{\underbrace{\sum (y_i - \bar{y})^2}_{\cancel{SSR}\ SST}}$$

→ 1 perfect expl.

→ 0 expl. nothing

$\boxed{\text{Frac of the var. in y expl by the Reg}}$

only in Single linear Reg. } $\Rightarrow (\text{Corr})^2 = R^2$

Y → corr

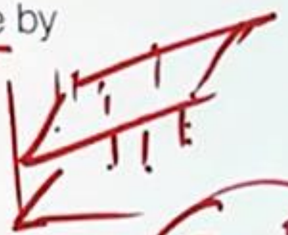
$$R^2 = \frac{\overbrace{SST - SSE}^{SSR}}{SST} = \frac{SSR}{SST}$$

Multiple Regression

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

$$\boxed{y = a + b_1 x_1 + \cdots + b_k x_k} \rightarrow \text{Hyper plane}$$

- 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.



Σe_i^2

- Very easily computed using most Statics of ML tools

y	x_1	x_2	x_3	x_4	x_5	x_6
mpg	cyl	disp	hp	wt	acc	yr
18	8	307	130	3504	12	70
15	8	350	165	3693	11.5	70
18	8	318	150	3436	11	70
16	8	304	150	3433	12	70
17	8	302	140	3449	10.5	70
15	8	429	198	4341	10	70
14	8	454	220	4354	9	70
14	8	440	215	4312	8.5	70
14	8	455	225	4425	10	70
15	8	390	190	3850	8.5	70

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

Data Source: StatLib (http://lib.stat.cmu.edu/datasets/)

Standard Error and Adjusted R^2

$\bar{R} \to \sum e_i^2$

$\bar{R} \to \sum (y_i - \bar{y})^2$

- Standard Error for Multiple regression

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

- Adjusted R^2
  - 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"

$y = a + b_1 x_1$

$y = a + b_1 x_1 + b_2 x_2$

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

$R^2 = 60\%$

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

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

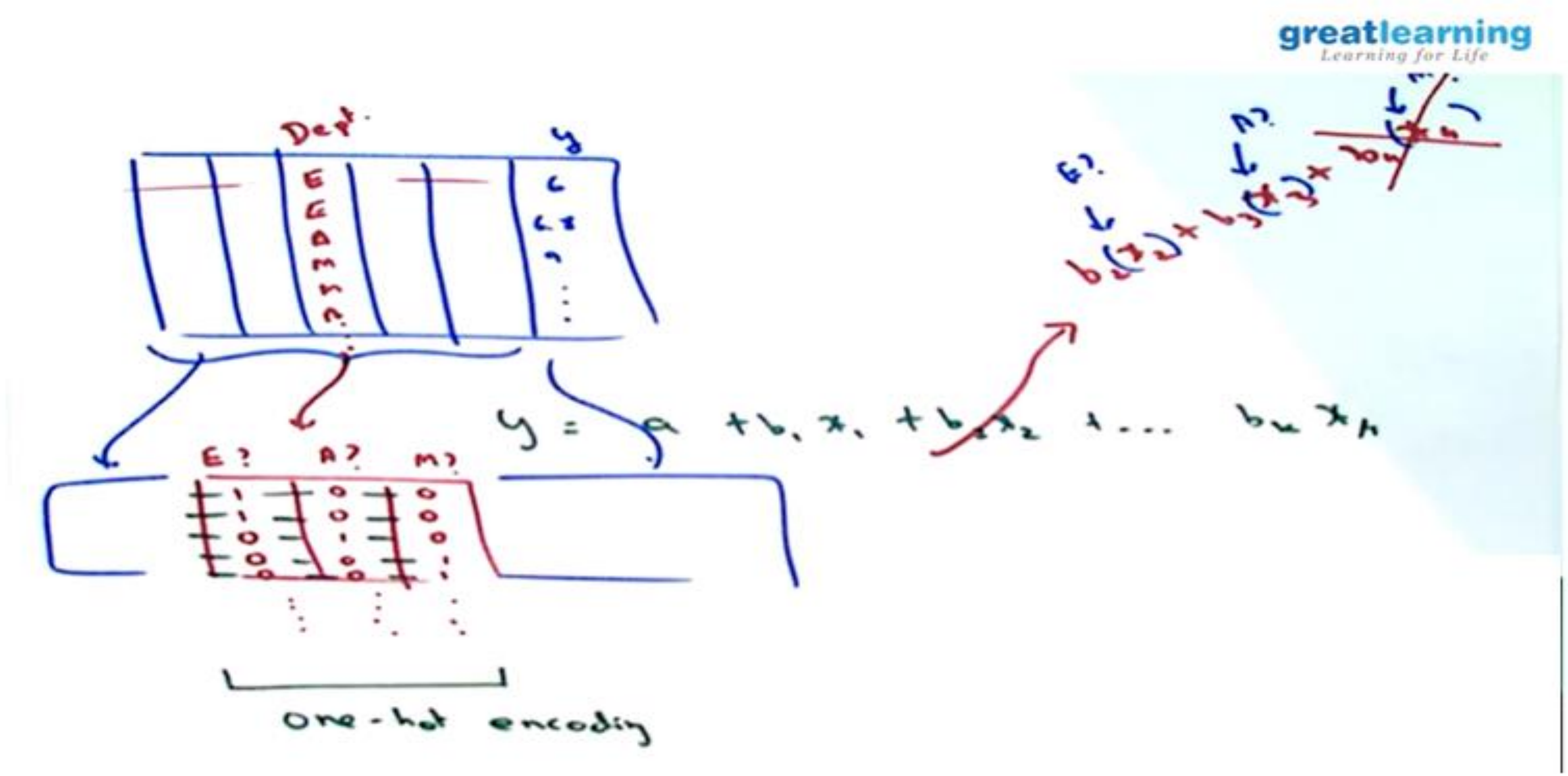
R^2 usually increase when var are added

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

but will go down if var is not adding <u>enough</u> value

One – Hot - Encoding

Categorical Independent variables in Regression



Dummy Var

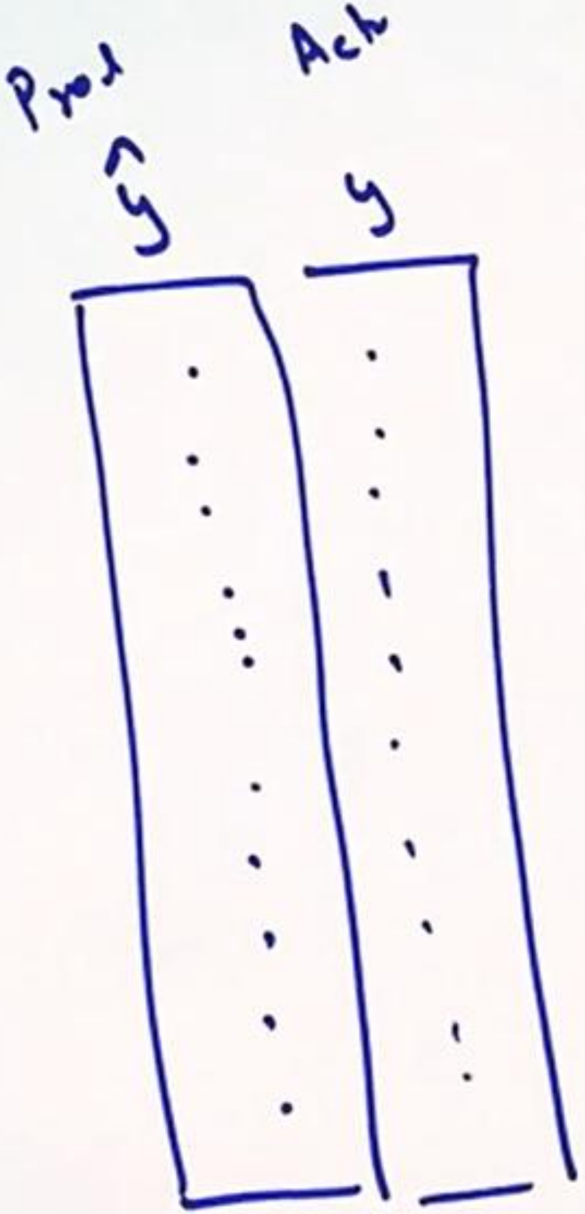
	Dept	
	E	
	E	
	M	
	A	

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

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

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

b_4 represents the diff from base (Dept = M)



MAE

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

RMSE

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

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

greatlearning
Learning for Life

Pros and Cons

(x_1, x_8)

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

$mpg = 46.317 - 0.0077 (wt)$

b_1

- 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