An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023

iHUB, DivyaSampark, IIT Roorkee

Ritvij Bharat Private Limited (RBPL)

Performance Evaluation

Regression Metrics

Presented By:
Shreyas Shukla

Fig Ref: Pixnio Data

Evaluating Regression

How do we decide if those predictions are any good?

Evaluating Regression

- Accuracy
- Recall.
- These sort of metrics aren't useful for regression problems, we need metrics designed for **continuous** values!

Evaluating Regression

- Mean Absolute Error
- Mean Squared Error
- Root Mean Square Error

Evaluating Regression

An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023

Mean Absolute Error (MAE)

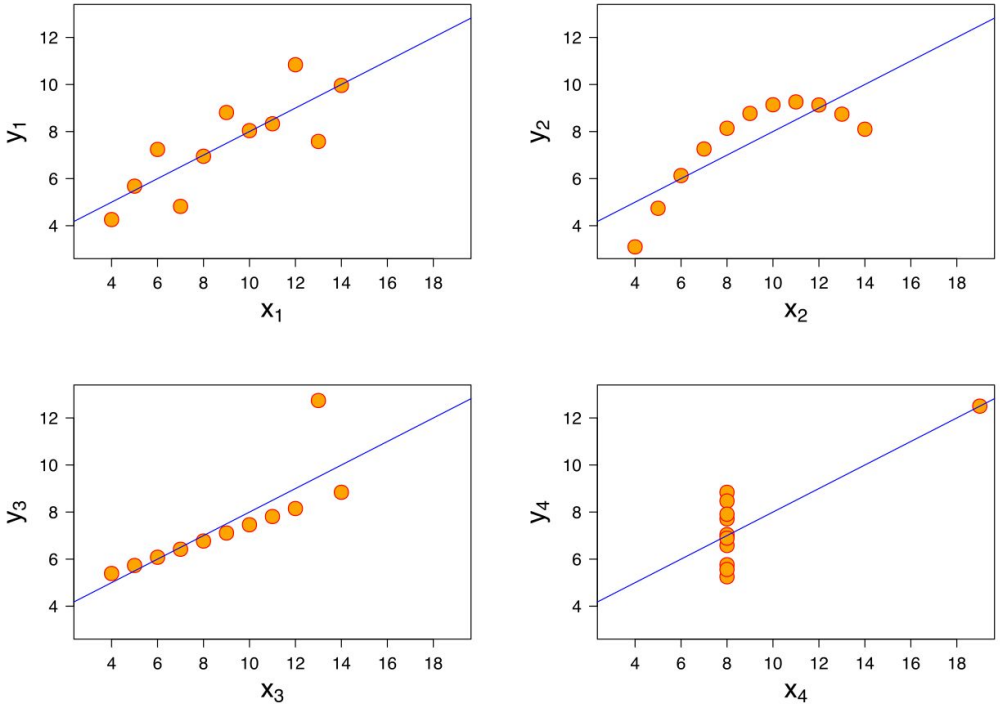
iHUB, DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited (RBPL)

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

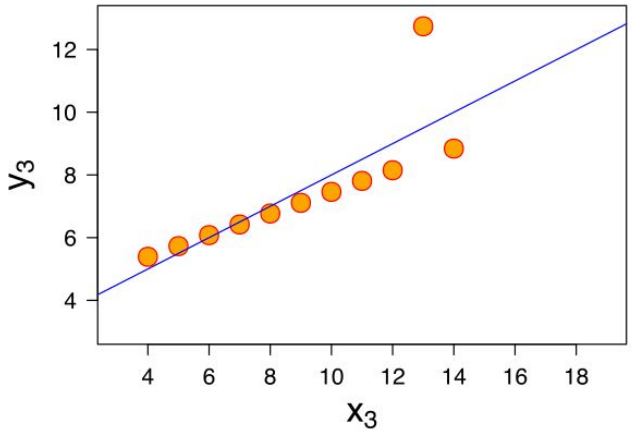
Fig Ref: Pierian Data

Evaluating Regression

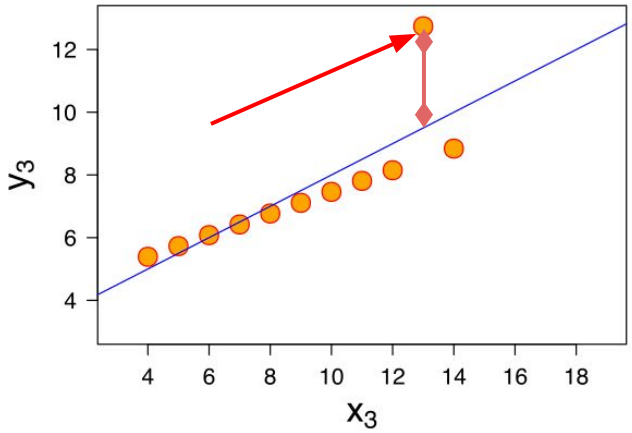
MAE won't punish large errors however.



Evaluating Regression



Evaluating Regression



Evaluating Regression

An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023
iHUB DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited (RBPL)

Mean Squared Error (MSE)

- Issue with MSE:
  - Different units than y.
  - It reports units of y squared!

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Fig Ref: Pierian Data

Evaluating Regression

An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023
iHUB DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited (RBPL)

- Root Mean Square Error (RMSE)
  - root of the mean of the squared errors.
  - Most popular (has same units as y)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}$$

Fig Ref: Pienaar Data

Machine Learning

An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023
and
Ritvij Bharat Private Limited (RBPL)
Presented By:
Shreyas Shukla

Compare your error metric to the average value of the label in your data set to try to get an intuition of its overall performance.

Domain knowledge also plays an important role here!

Machine Learning

Context of importance is also necessary to consider.
We may create a model to predict how much medication to give, in which case small fluctuations in RMSE may actually be very significant.

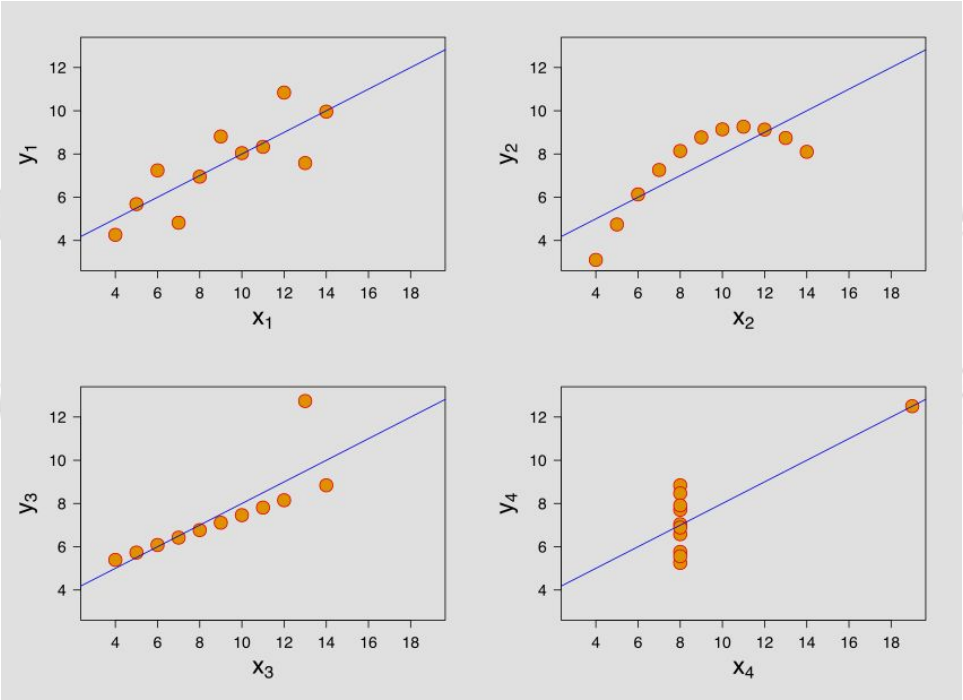
An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023

iHUB, DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited (RBPL)

Presented By:
Shreyas Shukla

Evaluating Residuals

Anscombe's quartet:



Clearly Linear Regression is not suitable

Linear Regression

An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023
iHUB DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited (RBPL)
Presented By:
Shreyas Shukla

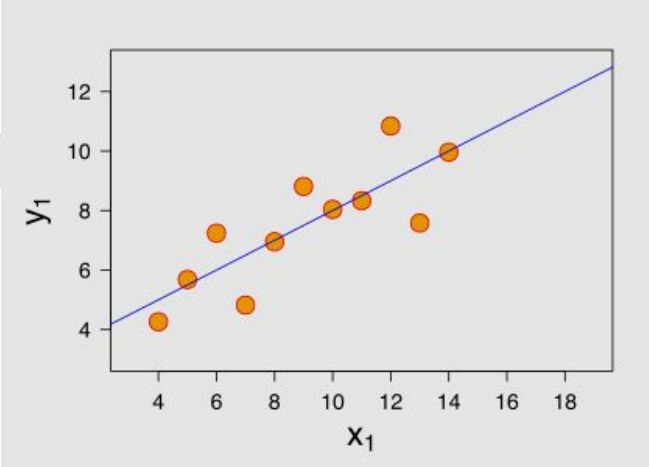
But how can we tell if we're dealing with more than one x feature?

Fig Ref: Pierian Data

Linear Regression

An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023
iHUB, DivyaSampark, IIT Roorkee
Ritvij Bi d (RBPL)

Consider an appropriate data set:



Linear Regression

An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023
iHUB, DivyaSampark, IIT Roorkee
and
Ritvij Bl (RBPL)

The residual errors.

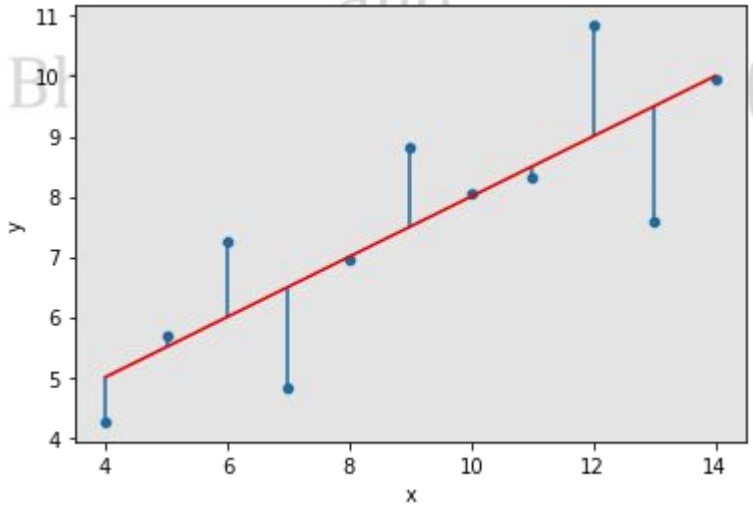


Fig Ref: Pierian Data

Linear Regression

An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023
iHUB, DivyaSampark, IIT Roorkee
and
Ritvij Bha RBPL)

The residual errors should be random and close to a normal distribution.

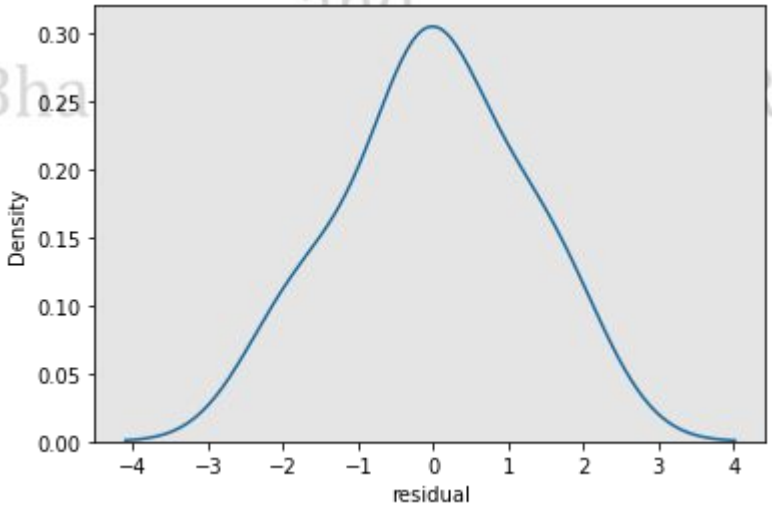


Fig Ref: Pierian Data

Linear Regression

An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023
iHUB DivyaSampark, IIT Roorkee
Ritvij Bha (RBPL)

Residual plot shows residual error vs. true y value.

There should be no clear line or curve.

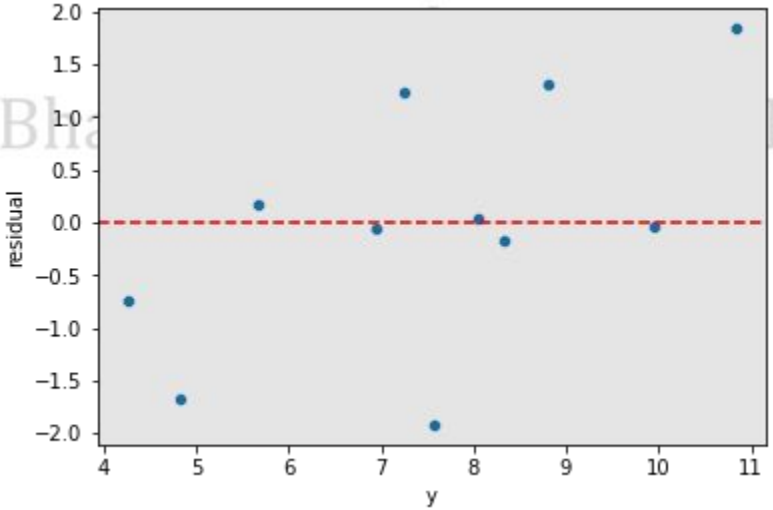
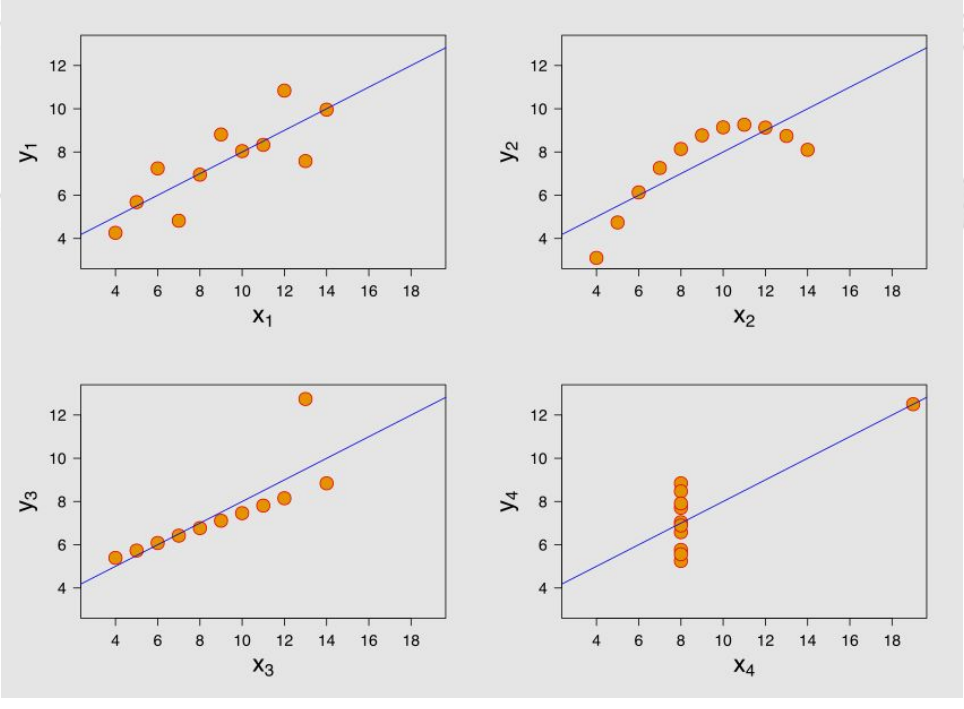


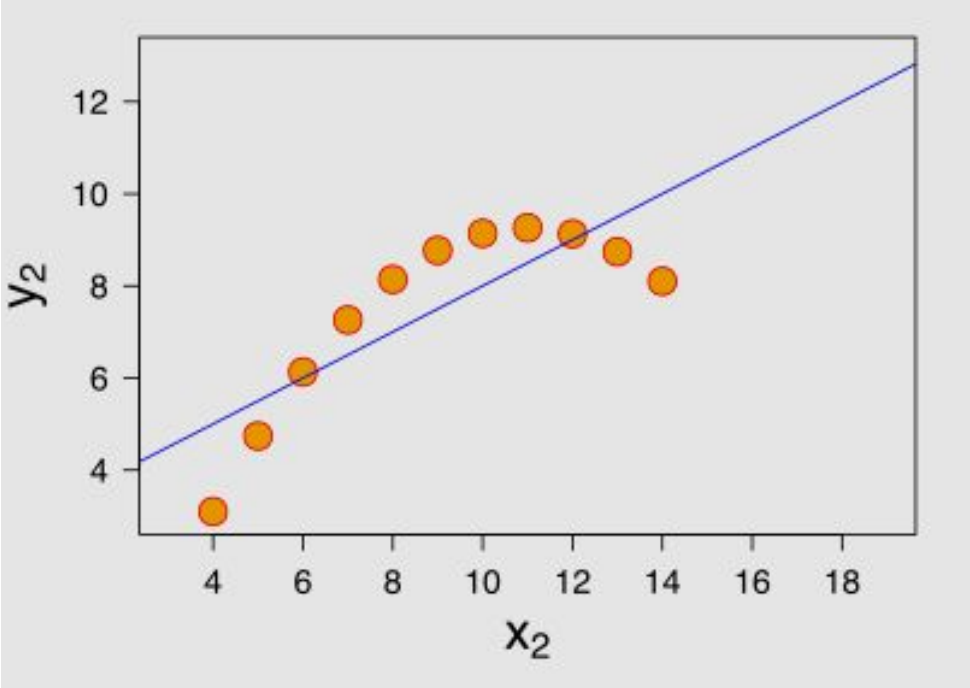
Fig Ref: Pierian Data

Linear Regression

Let’s see 2nd one



Linear Regression



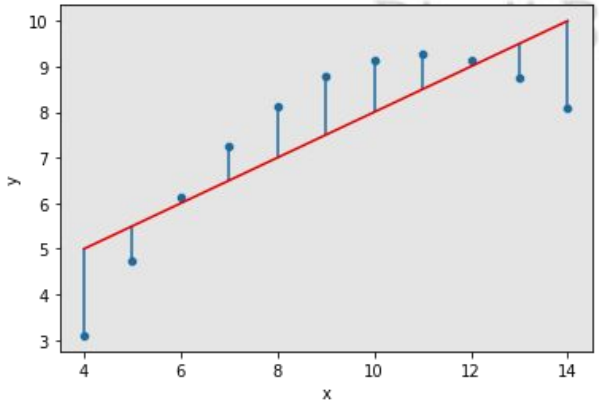
Linear Regression

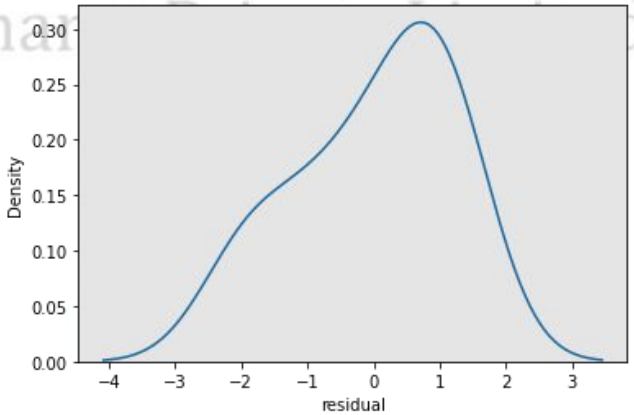
An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023

Residual plot showing a clear pattern, indicating Linear Regression no valid!

Especially the 3rd plot

iHUB, DivyaSampark, IIT Roorkee
and





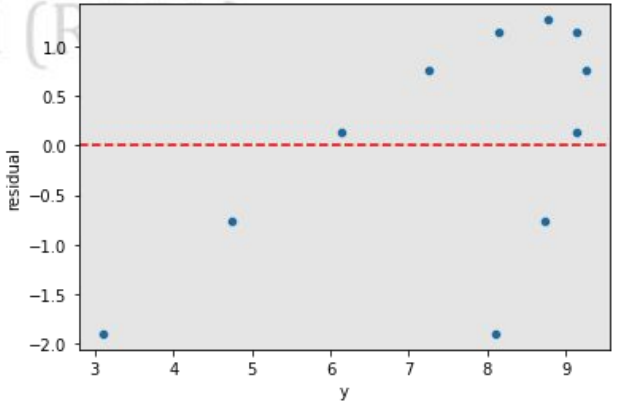
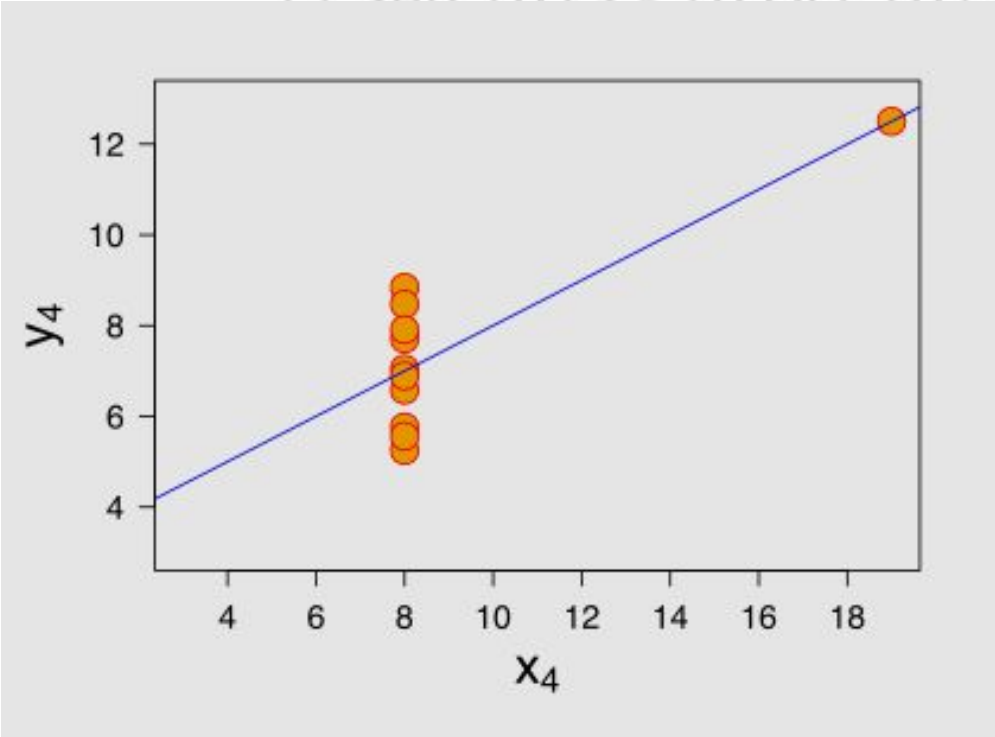


Fig Ref: Pierian Data

Linear Regression

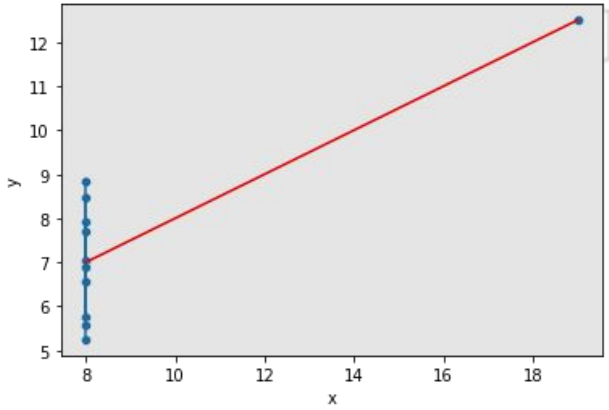


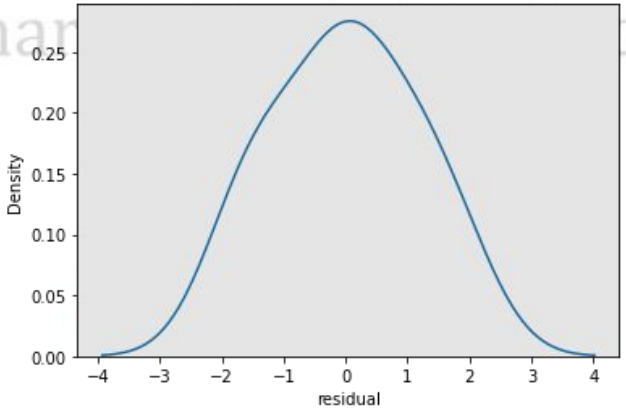
Linear Regression

An Introduction to Machine Learning with Python Programming
12 Sep 2023 - 20 Oct 2023

Residual plot showing a clear pattern, indicating Linear Regression not valid!

iHUB, DivyaSampark, IIT Roorkee
and





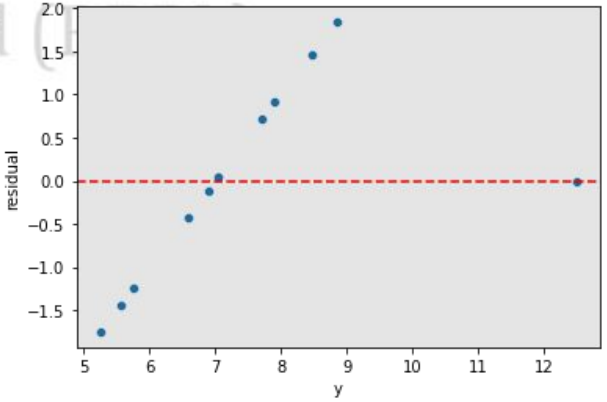


Fig Ref: Pierian Data

Let’s create in Python

iHUB, DivyaSampark, IIT Roorkee
and
Ritvij Bharat Private Limited (RBPL)

Presented By:
Shreyas Shukla

Fig Ref: Pienan Data