

Nama: Dilara Kynta Putri Rafliita
NIM: 1103204059
Kelas: Machine Learning TK44GAB4

Understanding 3 Link StatQuest (Youtube: Josh Starmer)

1) Principal Component Analysis (PCA) Step-by-step

| | Mouse 1 | Mouse 2 | Mouse 3 | Mouse 4 | Mouse 5 | Mouse 6 |
|--------|---------|---------|---------|---------|---------|---------|
| Gene 1 | 10 | 11 | 8 | 3 | 2 | 1 |
| Gene 2 | 6 | 4 | 5 | 3 | 2.8 | 1 |
| Gene 3 | 12 | 9 | 10 | 2.5 | 1.3 | 2 |
| Gene 4 | 5 | 20 | 6 | 2 | 18 | 19 |

If we measured 4 genes per mouse, we would not be able to draw a 4-dimensional graph of the data...
but that doesn't stop us from doing the PCA math, which doesn't care if we can draw

StatQuest: Principal Component Analysis (PCA), Step-by-Step

StatQuest with Josh Starmer

2.4M views 5 years ago High Throughput Sequencing

PCA used to reduce the dimensionality of data while preserving as much variance as possible. This is a step by step guide on how PCA works:

- Conceptual motivation for PCA

| | Student 1 | Student 2 | Student 3 | Student 4 | ... |
|---------|-----------|-----------|-----------|-----------|-----|
| Math | 95 | 88 | 93 | 75 | ... |
| Reading | 96 | 79 | 98 | 81 | ... |

For example, the samples could be students in high school and the variables could be test scores in math and reading...

StatQuest: Principal Component Analysis (PCA), Step-by-Step

StatQuest with Josh Starmer

1.01M views 5 years ago

- Worked out for 2-dimensional data

https://www.youtube.com/watch?v=FgakZw6K1QQ

StatQuest: Principal Component Analysis (PCA), Step-by-Step

StatQuest... 1.01M... Join Subscribe 50K Share

2,453,797 views Apr 3, 2018 High Throughput Sequencing

monday.com Make gantt's great again

Sign Up Ad · monday.com Start now

All From StatQuest with Josh Starmer Related

Machine Learning Algorithms Explained

Understand Where To Start When Learning The Basics Of Machine Learning Concepts.

Ad · Dataiku Download

3 Practical Tips... StatQuest: PCA - Practical Tips

StatQuest with Josh Starmer 131K views · 5 years ago

- Finding PC 1

https://www.youtube.com/watch?v=FgakZw6K1QQ&t=351s

StatQuest: Principal Component Analysis (PCA), Step-by-Step

StatQuest... 1.01M... Join Subscribe 50K Share

2.4M views 5 years ago High Throughput Sequencing

Tania Cremenko NEW VIDEO EVERY WEEK

Online webinar Ad · sgvirtualassistant.com Join now

All From StatQuest with Josh Starmer Related

12 Weeks Program by MIT IDSS

Learn Machine Learning, Deep Learning, NLP, Recommendation Systems, and more. Apply Now.

Ad · Great Learning Learn more

3 Practical Tips... StatQuest: PCA - Practical Tips

StatQuest with Josh Starmer 131K views · 5 years ago

PCA Main StatQuest: PCA main ideas in only 5 minutes!!!

- Finding PC 2

https://www.youtube.com/watch?v=FgakZw6K1QQ&t=351s

YouTube

Search

Gene 2

Gene 1

...or it can try to find the line that **maximizes** the distances from the projected points to the origin.

line that minimizes those distances, or it can try to find the line that maximizes

StatQuest: Principal Component Analysis (PCA), Step-by-Step

StatQuest... 1.01M... Join Subscribe 50K Share

2.4M views 5 years ago High Throughput Sequencing

Online webinar Ad · sgbvirtualassistant.com Join now

12 Weeks Program by MIT IDSS Learn Machine Learning, Deep Learning, NLP, Recommendation Systems, and more. Apply Now. Ad · Great Learning Learn more

3 Practical Tips... StatQuest: PCA - Practical Tips StatQuest with Josh Starmer 131K views · 5 years ago

PCA Main StatQuest: PCA main ideas in only 5 minutes!!!

- Drawing the PCA graph

https://www.youtube.com/watch?v=FgakZw6K1QQ&t=351s

YouTube

Search

PC2

PC1

...Sample 2 goes here.

example, these projected points correspond to sample 6, so sample 6 goes here.

Sample

StatQuest: Principal Component Analysis (PCA), Step-by-Step

StatQuest... 1.01M... Join Subscribe 50K Share

2.4M views 5 years ago High Throughput Sequencing

Online webinar Ad · sgbvirtualassistant.com Join now

12 Weeks Program by MIT IDSS Learn Machine Learning, Deep Learning, NLP, Recommendation Systems, and more. Apply Now. Ad · Great Learning Learn more

3 Practical Tips... StatQuest: PCA - Practical Tips StatQuest with Josh Starmer 131K views · 5 years ago

PCA Main StatQuest: PCA main ideas in only 5 minutes!!!

- calculate percentation for each PC and scree plot

For the sake of the example, imagine that the Variation for **PC1 = 15**, and the variation for **PC2 = 3**.

That means that the total variation around both PCs is **15 + 3 = 18**...

...and that means PC1 accounts for **15 / 18 = 0.83 = 83%** of the total variation around the PCs.

And that means PC1 accounts for 15 divided by 18 equals 0.83 or 83% of the total variation

StatQuest: Principal Component Analysis (PCA), Step-by-Step

2.4M views · 5 years ago · High Throughput Sequencing

SS(distances for PC1) = Variation for PC1

$$\frac{SS(\text{distances for PC1})}{n - 1}$$

SS(distances for PC2) = Variation for PC2

$$\frac{SS(\text{distances for PC2})}{n - 1}$$

PC2

PC1 (83%)

Online webinar
 Ad · sgbvirtualassistant.com

12 Weeks Program by MIT IDSS
 Learn Machine Learning, Deep Learning, NLP, Recommendation Systems, and more. Apply Now.

3 Practical Tips...
 ...for PCA

StatQuest: PCA - Practical Tips
 StatQuest with Josh Starmer
 131K views · 5 years ago

PCA Main

- PCA worked for 3 dimensional data

Lastly, we find PC3, the best fitting line that goes through the origin and is perpendicular to PC1 and PC2...

line that goes through the origin and is perpendicular PC1 and PC2. If we had more

StatQuest: Principal Component Analysis (PCA), Step-by-Step

https://www.youtube.com 5 years ago High Throughput Sequencing

Gene 2

Gene 3

Gene 1

PC1

PC2

PC3

Online webinar
 Ad · sgbvirtualassistant.com

12 Weeks Program by MIT IDSS
 Learn Machine Learning, Deep Learning, NLP, Recommendation Systems, and more. Apply Now.

3 Practical Tips...
 ...for PCA

StatQuest: PCA - Practical Tips
 StatQuest with Josh Starmer
 131K views · 5 years ago

PCA Main

2) ROC and AUC explained

The screenshot shows a YouTube video player. The video title is "ROC and AUC, Clearly Explained!". The channel is "StatQuest with Josh Starmer" with 1.01M subscribers. The video has 1.2M views and was uploaded 4 years ago. The video content shows a scatter plot with "Is Obese" on the y-axis and "Weight" on the x-axis. There are red dots for "Is Not Obese" and blue dots for "Is Obese". A vertical line is drawn at a certain weight, and a point on the "Is Not Obese" line is labeled "It must be Mighty Mouse and just full of muscles." The video player interface includes a search bar, a subscribe button, and a share button. The browser tabs show "ROC and AUC, Cle...", "Submit Tugas Machine...", "Car Evaluation Data Set...", "kNN Classifier Tutorial...", and "UCL_Breast Cancer Wit...". The browser address bar shows "https://www.youtube.com/watch?v=4jRBRDbJemM". The system tray at the bottom shows the date "11/10/2023" and time "10:15".

- Classifying samples with logistic regression

The screenshot shows a YouTube video player. The video title is "ROC and AUC, Clearly Explained!". The channel is "StatQuest with Josh Starmer" with 1.01M subscribers. The video has 1.2M views and was uploaded 4 years ago. The video content shows a sigmoid curve representing the probability of obesity vs weight. A horizontal dashed line at 0.5 is labeled "One way to classify mice is to set a threshold at 0.5...". The video player interface includes a search bar, a subscribe button, and a share button. The browser tabs show "ROC and AUC, Cle...", "Submit Tugas Machine...", "Car Evaluation Data Set...", "kNN Classifier Tutorial...", and "UCL_Breast Cancer Wit...". The browser address bar shows "https://www.youtube.com/watch?v=4jRBRDbJemM". The system tray at the bottom shows the date "11/10/2023" and time "10:15".

- Creating a confusion matrices for different thresholds

https://www.youtube.com/watch?v=4jRBRDbJemM

YouTube

Search

One way to classify mice is to set a threshold at 0.5...

The probability of obesity

Weight

ROC and AUC, Clearly Explained!

StatQuest... 1.01M...

Join Subscribe

30K

Share

1.2M views 4 years ago #66DaysOfData

DATASTAX

DataStax - Vector Database Ad · datastax.com Visit site

All From StatQuest with Josh Star... Related

Download 8 million+ graphics Premium Crafting Fonts, Graphics & More Ad · Creative Fabrica Download

Ridge Regression... Regularization Part 1: Ridge (L2) Regression StatQuest with Josh Starmer 903K views · 5 years ago

PCA... StatQuest: Principal Component Analysis (PCA) StatQuest

- ROC is an alternative to tons of confusion matrices

https://www.youtube.com/watch?v=4jRBRDbJemM

YouTube

Search

So if the red ROC curve represented Logistic Regression and the blue ROC curve represented a Random Forest, you would use the Logistic Regression.

True Positive Rate (Sensitivity)

False Positive Rate (1 - Specificity)

ROC and AUC, Clearly Explained!

StatQuest... 1.01M...

Join Subscribe

30K

Share

1.2M views 4 years ago #66DaysOfData

DATASTAX

DataStax - Vector Database Ad · datastax.com Visit site

All From StatQuest with Josh Star... Related

Download 8 million+ graphics Premium Crafting Fonts, Graphics & More Ad · Creative Fabrica Download

Ridge Regression... Regularization Part 1: Ridge (L2) Regression StatQuest with Josh Starmer 903K views · 5 years ago

PCA... StatQuest: Principal Component Analysis (PCA) StatQuest

- AUC to compare different models

True Positive Rate = Sensitivity = $\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$

| | | Actual | |
|-----------|--------------|-----------------|-----------------|
| | | Is Obese | Is Not Obese |
| Predicted | Is Obese | True Positives | False Positives |
| | Is Not Obese | False Negatives | True Negatives |

The **True Positive Rate** tells you what proportion of **obese** samples were correctly classified.

ROC and AUC, Clearly Explained!

StatQuest... 1.01M... Join Subscribe 30K Share ...

1.2M views 4 years ago #66DaysOfData

DATASTAX
DataStax - Vector Database
Ad · datastax.com Visit site

All From StatQuest with Josh Starmer Related

Download 8 million+ graphics
Premium Crafting Fonts, Graphics & More
Ad · Creative Fabrica Download

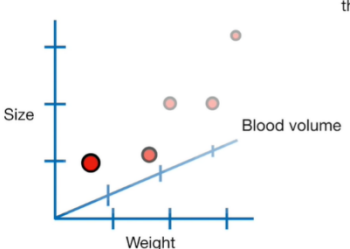
Ridge Regression... Regularization Part 1: Ridge (L2) Regression
StatQuest with Josh Starmer 903K views · 5 years ago

PCA... StatQuest: Principal Component

3) Logistic Regression

Logistic regression is commonly used in machine learning for classification problems.

Multiple regression did the same things that normal regression did...



Size

Weight

Blood volume

StatQuest: Logistic Regression

StatQuest... 1.01M... Join Subscribe 33K Share ...

Learn with Videos & Online Classes
Start learning

Video Notebook
Ad · videonotebook.com Sign up

All From StatQuest with Josh Starmer Related

Machine Learning Algorithms Explained
Understand Where To Start When Learning The Basics Of Machine Learning Concepts.
Ad · Dataiku Download

Logistic Regression Details Pt 1...
...coeffic 19:02

Logistic Regression Details Pt1: Coefficients
StatQuest with Josh Starmer 776K views · 5 years ago

- Review of linear regression

Click to go back (Alt+Left arrow), hold to see history

https://www.youtube.com/watch?v=yYKR4sgzi8

YouTube

Search

Size

Weight

...and with that line, we could do a lot of things:

1) Calculate R^2 and determine if **weight** and **size** are correlated. Large values imply a large effect.

StatQuest: Logistic Regression

StatQuest... 1.01M...

Join Subscribe

33K

Share

1.8M views 5 years ago Logistic Regression

Machine Learning Algorithms Explained

Understand Where To Start When Learning The Basics Of Machine Learning Concepts.

Ad · Dataiku

Download

Logistic Regression Details Pt 1...

Logistic Regression Details Pt1: Coefficients

StatQuest with Josh Starmer

776K views · 5 years ago

p-values: What they are and

https://www.youtube.com/watch?v=yYKR4sgzi8

YouTube

Search

Size

WT Single KO Double KO

We also talked about how we can use discrete measurements, like **genotype**, to predict **size**.

StatQuest: Logistic Regression

StatQuest... 1.01M...

Join Subscribe

33K

Share

1.8M views 5 years ago Logistic Regression

Learn with Videos & Online Classes

Start learning

Video Notebook

Ad · videonotebook.com

Sign up

Machine Learning Algorithms Explained

Understand Where To Start When Learning The Basics Of Machine Learning Concepts.

Ad · Dataiku

Download

Logistic Regression Details Pt 1...

Logistic Regression Details Pt1: Coefficients

StatQuest with Josh Starmer

776K views · 5 years ago

Comparing the simple model to the complicated one tells us if we need to measure **weight** and **blood volume** to accurately predict **size** or if we get away with just **weight**.

StatQuest: Logistic Regression

StatQuest... 1.01M... Join Subscribe 33K Share

1.8M views 5 years ago Logistic Regression

Learn with Videos & Online Classes

Video Notebook Ad · videonotebook.com Sign up

Machine Learning Algorithms Explained

Understand Where To Start When Learning The Basics Of Machine Learning Concepts.

Ad · Dataiku Download

Logistic Regression Details Pt 1... coefficient 19:02

Logistic Regression Details Pt1: Coefficients

StatQuest with Josh Starmer 776K views · 5 years ago

- Logistic regression review

...also, instead of fitting a line to the data, logistic regression fits an "S" shaped "logistic function".

StatQuest: Logistic Regression

StatQuest... 1.01M... Join Subscribe 33K Share

1.8M views 5 years ago Logistic Regression

Learn with Videos & Online Classes

Video Notebook Ad · videonotebook.com Sign up

Machine Learning Algorithms Explained

Understand Where To Start When Learning The Basics Of Machine Learning Concepts.

Ad · Dataiku Download

Logistic Regression Details Pt 1... coefficient 19:02

Logistic Regression Details Pt1: Coefficients

StatQuest with Josh Starmer 776K views · 5 years ago

Step a step how to perform logistic regression:

- Data collection: gather a dataset that contains predictor variables (independent variables and dependent variables)
- Data preprocessing
- Split the data: divide the dataset into two parts, a training set and a testing (or validation) set.
- Model building: fit a logistic regression model to the training data.
- Model training: use the training data to estimate the coefficients value using an optimization algorithm

- Model evaluation: assess the performance of logistic regression model using the testing or validation data
- Interpretation: examine the estimated coefficients to understand the impact of each independent variable
- Predictions
- Deployment