



UNIT 4

LOGISTIC REGRESSION

Content Curated by Pollux M. Rey

FOR THIS UNIT...

01

What is Logistic Regression?

02

How does it work?

03

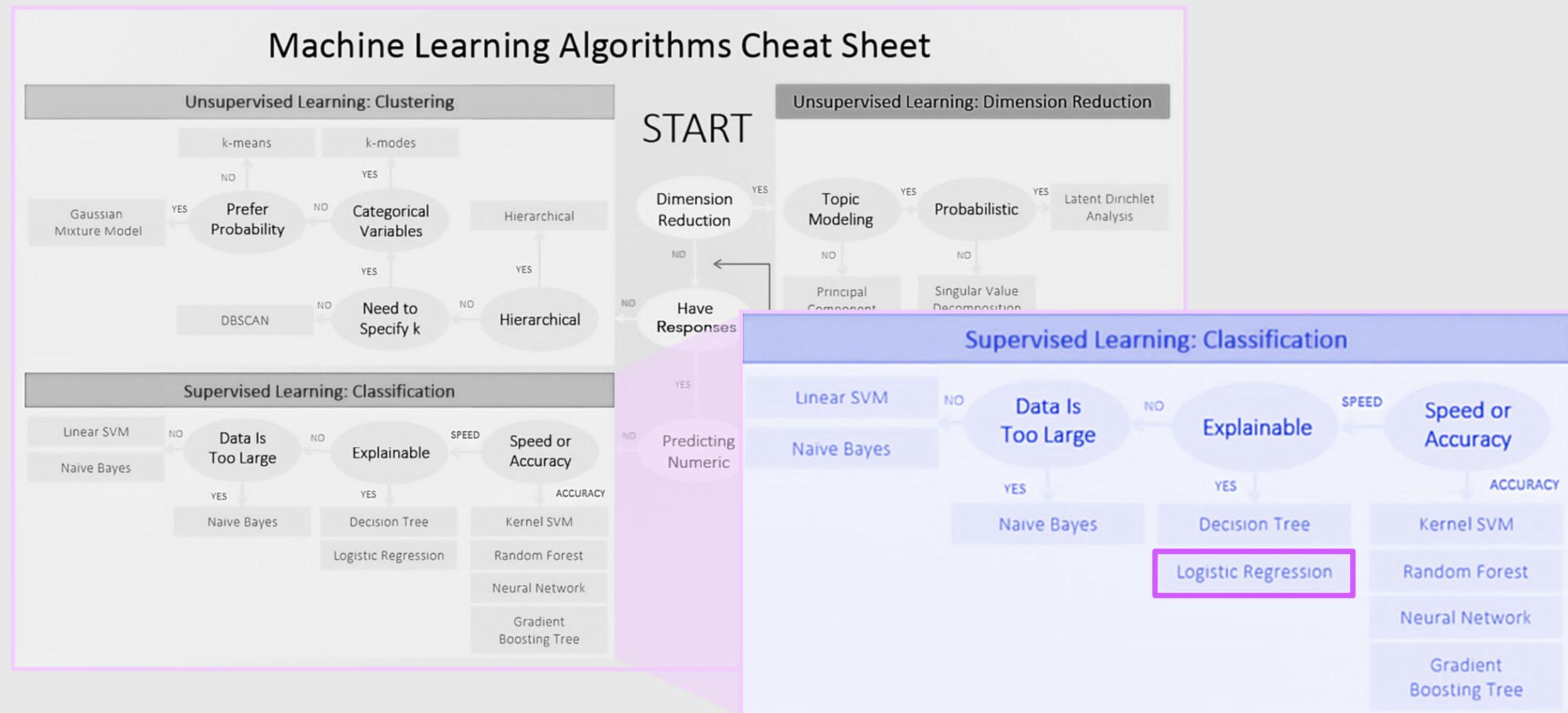
Performance Metrics

04

Lab Exercise

Content Curated by Pollux M. Rey

Logistic Regression



Content Curated by Pollux M. Rev

<https://blogs.sas.com/content/subconsciousmusings/2020/12/09/machine-learning-algorithm-use>

Logistic Regression

A supervised learning algorithm used for binary classification by predicting the probability of an observation belonging to a class based on its features.

It can also be extended to multiple categories.

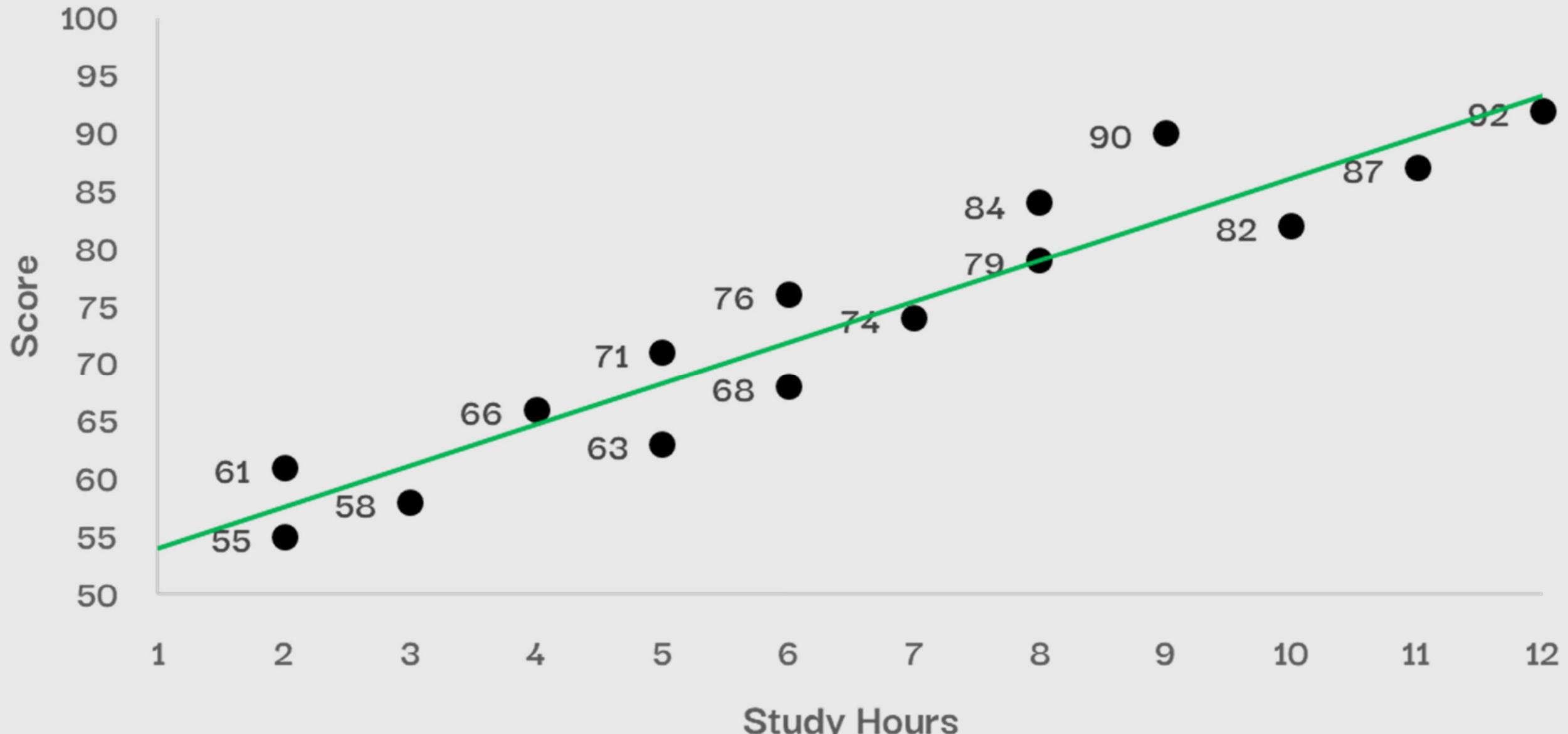
Linear Regression

| Study Hours (x) | Test Score (y) |
|--------------------|-------------------|
| 2 | 55 |
| 2 | 61 |
| 3 | 58 |
| 4 | 66 |
| 5 | 63 |
| 5 | 71 |
| 6 | 68 |
| 6 | 76 |
| 7 | 74 |
| 8 | 79 |
| 8 | 84 |
| 9 | 90 |
| 10 | 82 |
| 11 | 87 |
| 12 | 92 |

Logistic Regression

| Study Hours (x) | Passed? (y) |
|--------------------|----------------|
| 2 | No |
| 2 | No |
| 3 | No |
| 4 | No |
| 5 | No |
| 5 | No |
| 6 | No |
| 6 | Yes |
| 7 | No |
| 8 | Yes |
| 8 | Yes |
| 9 | Yes |
| 10 | Yes |
| 11 | Yes |
| 12 | Yes |

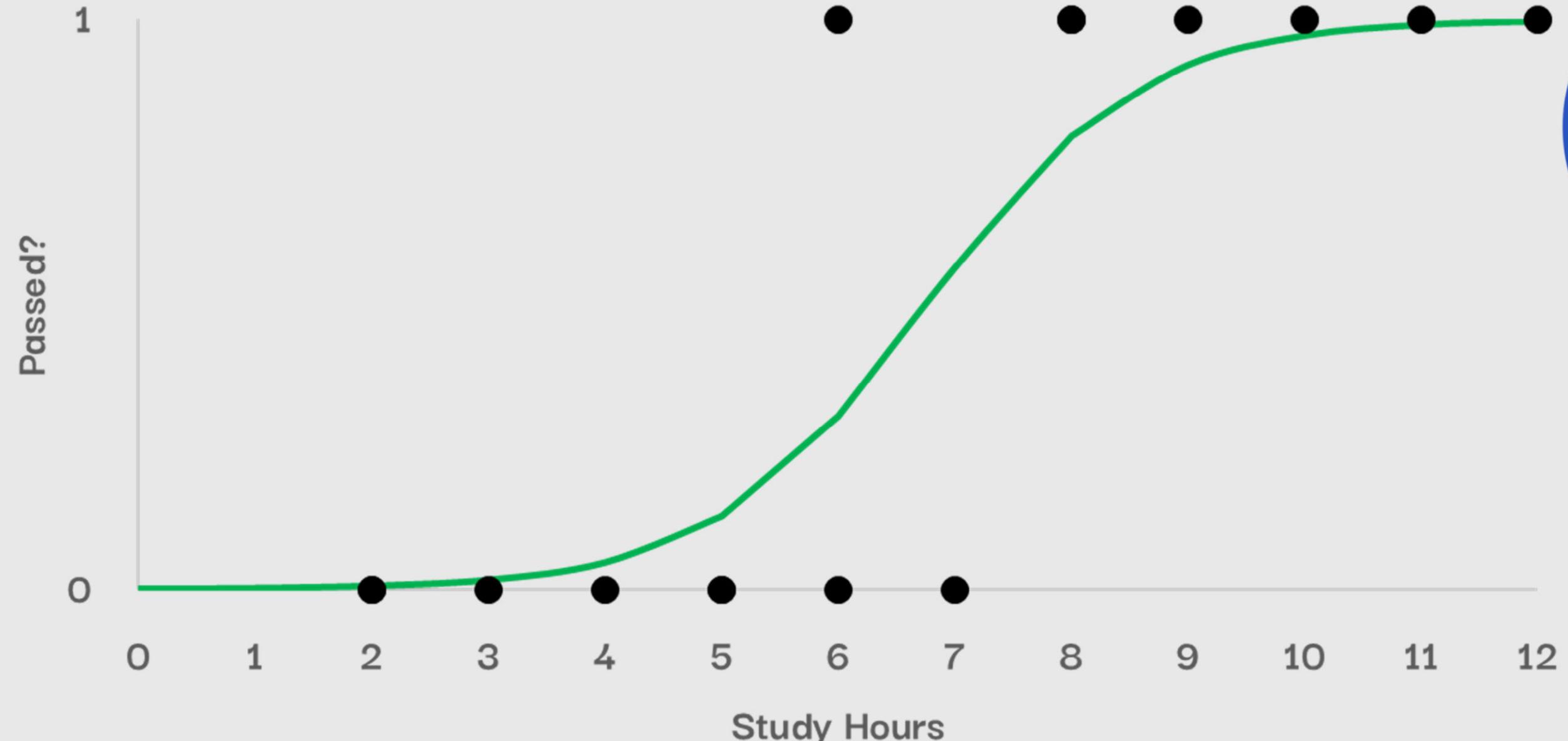
Linear Regression



Content Curated by Pollux M. Rey

<https://mlu-explain.github.io/logistic-regression/>
<https://www.datacamp.com/tutorial/understanding-logistic-regression-python>

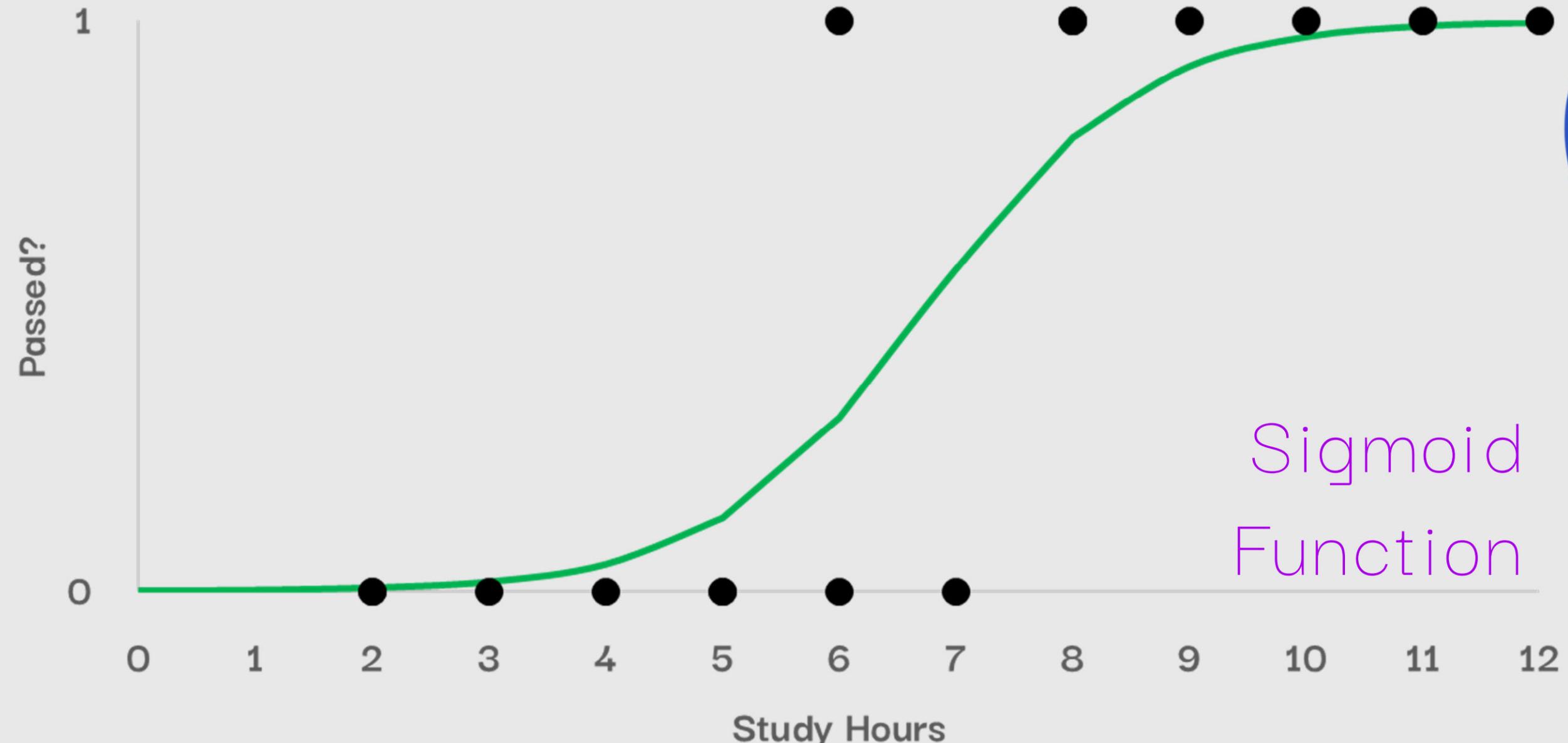
Logistic Regression



Content Curated by Pollux M. Rey

<https://mlu-explain.github.io/logistic-regression/>
<https://www.datacamp.com/tutorial/understanding-logistic-regression-python>

Logistic Regression



Logistic Regression – Sigmoid Function

$$p = \frac{1}{1 + e^{-y}}$$

where

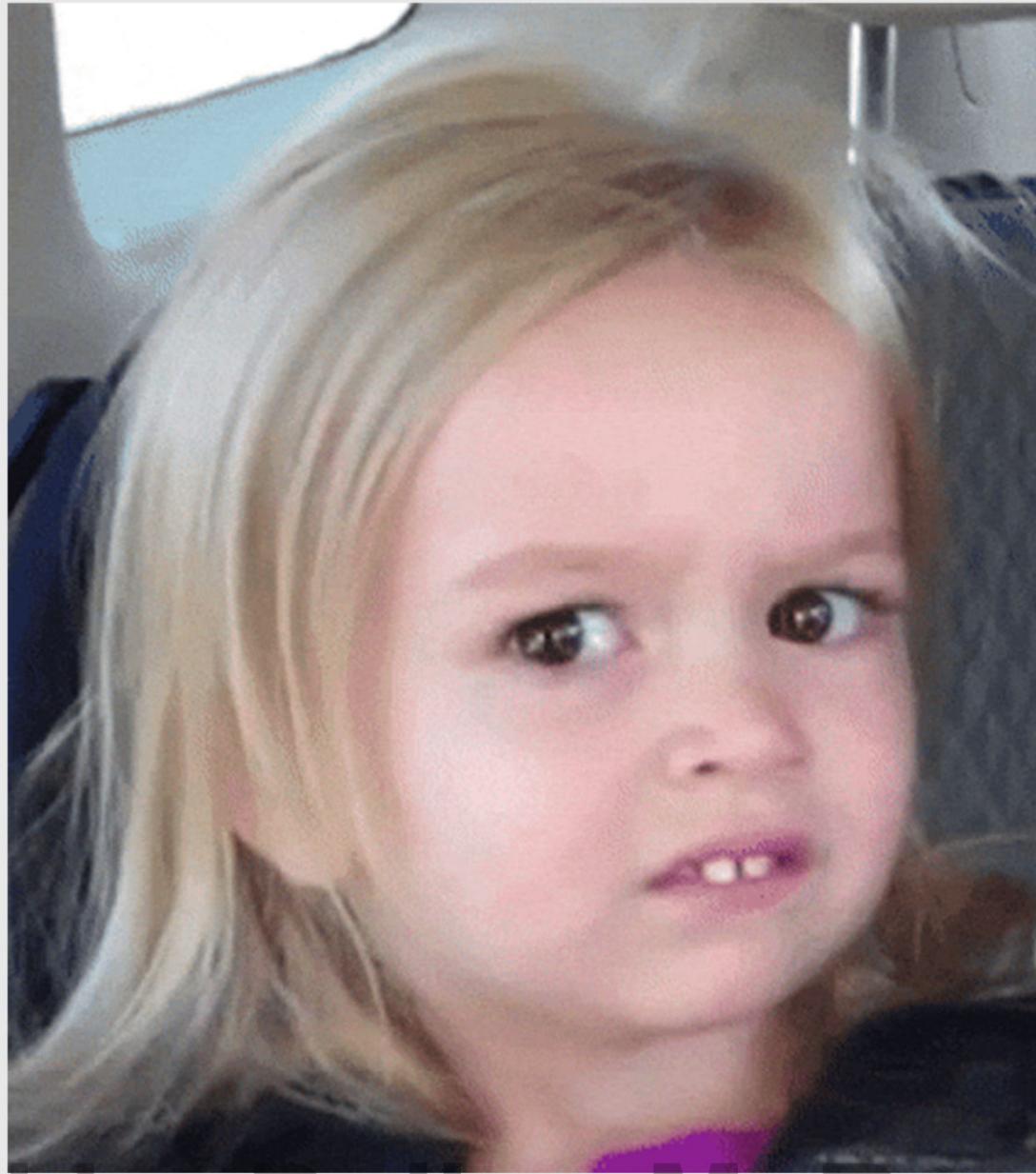
$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

Logistic Regression – Sigmoid Function



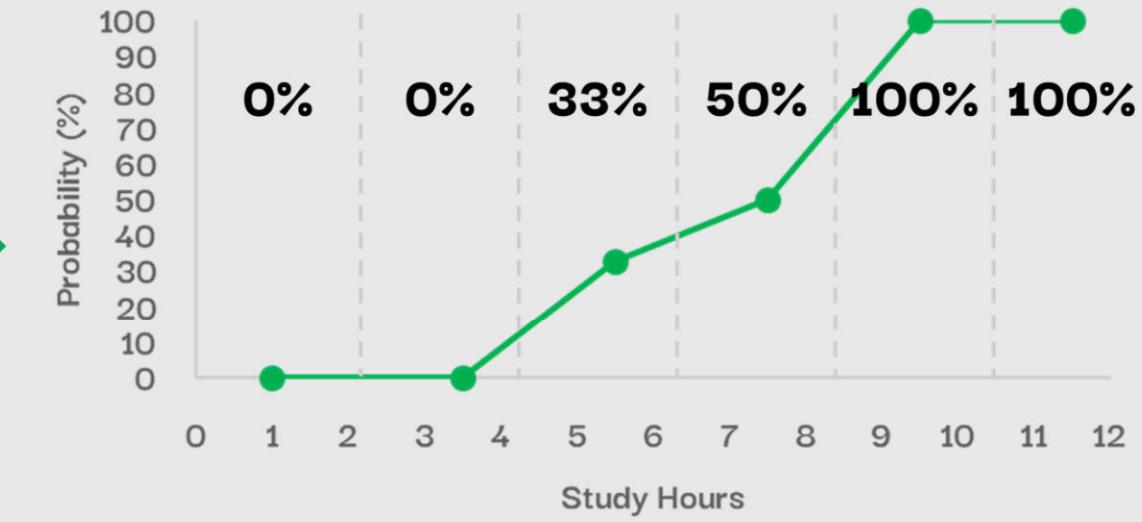
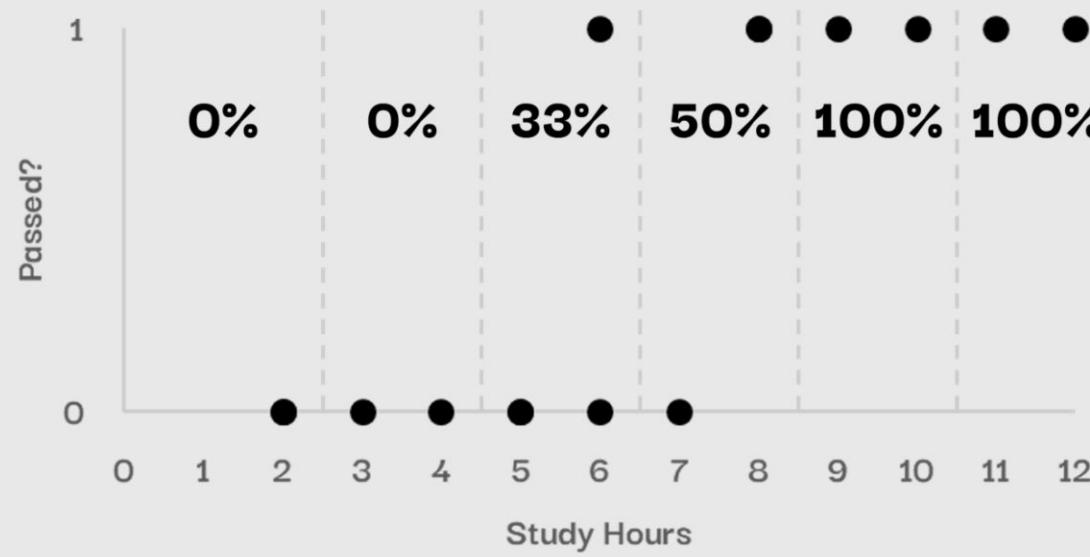
Content Curated by Pollux M. Rey

Logistic Regression – Sigmoid Function

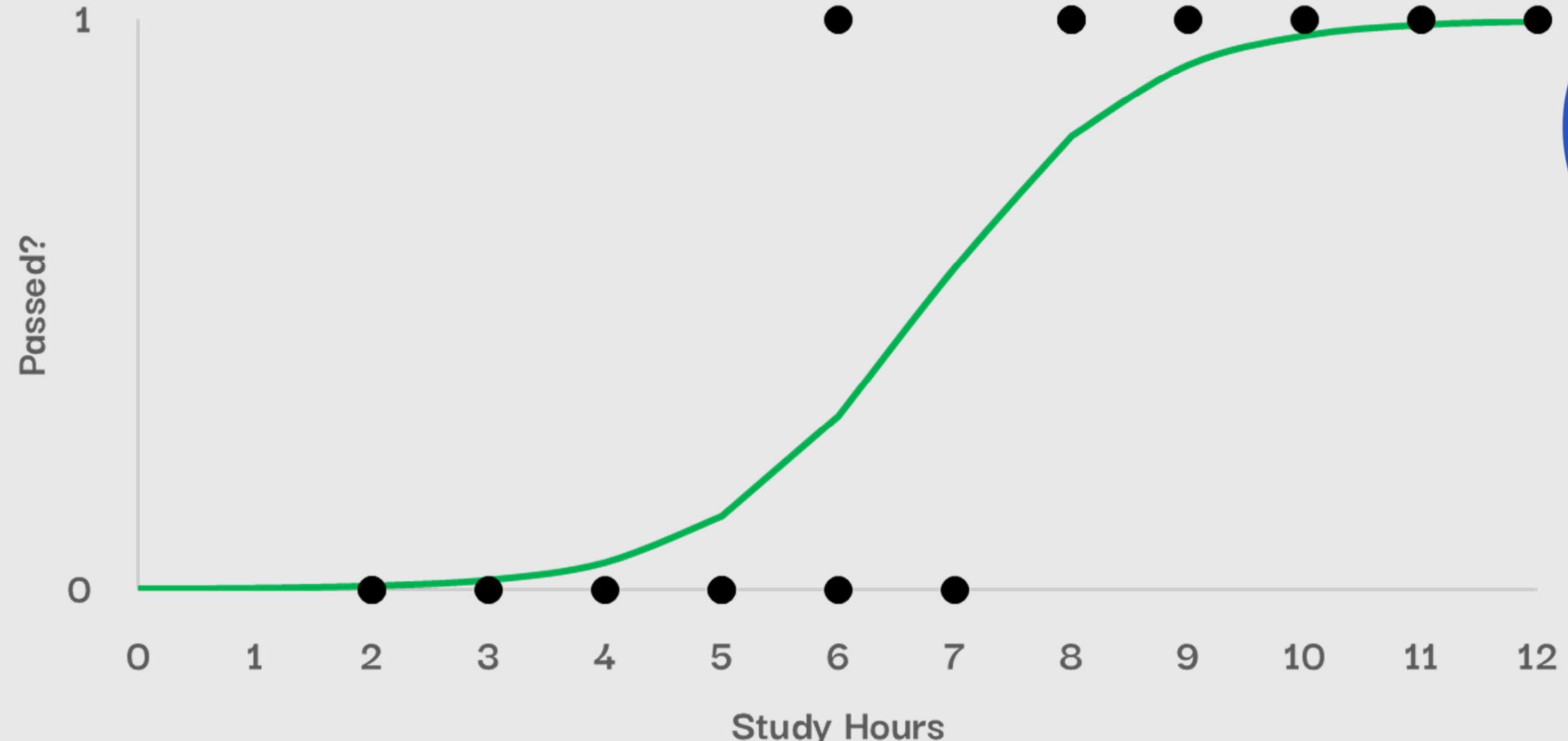


Content Curated by Pollux M. Rey

Logistic Regression



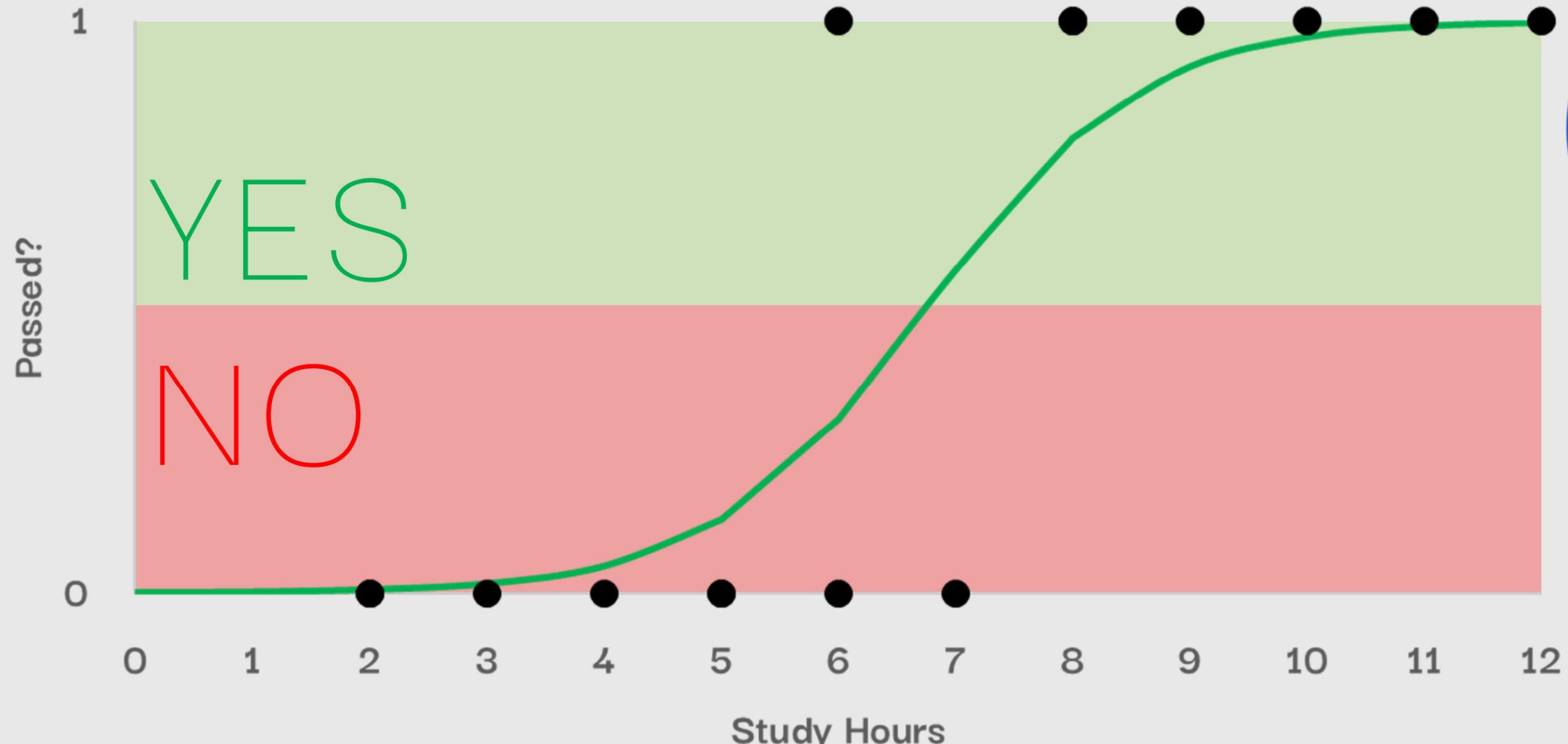
Logistic Regression



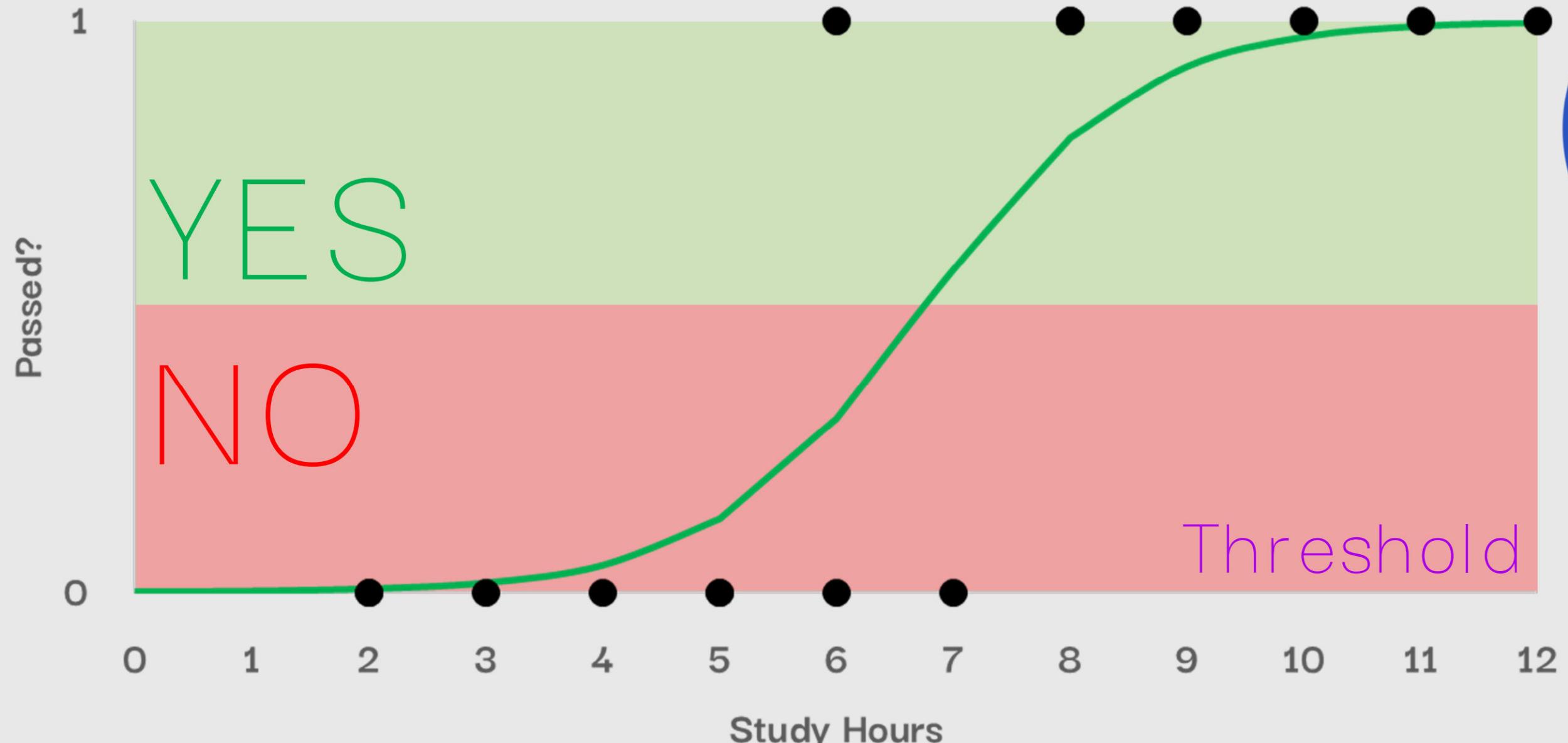
Content Curated by Pollux M. Rev

<https://www.youtube.com/watch?v=yhogDBEa0uQ>

Logistic Regression



Logistic Regression



Logistic Regression

[MUSIC PLAYING]



Content Curated by Pollux M. Rev

<https://www.youtube.com/watch?v=72AHKztZN44>

Logistic Regression Evaluation Metrics

| Study Hours (x) | Passed? (y) | Predictions (y') |
|--------------------|----------------|---------------------|
| 5 | No | No |
| 1 | No | No |
| 2 | Yes | No |
| 3 | No | No |
| 3 | No | No |
| 4 | No | No |
| 6 | Yes | Yes |
| 7 | Yes | Yes |
| 8 | No | Yes |
| 9 | Yes | Yes |
| 10 | Yes | Yes |
| 11 | Yes | Yes |
| 11 | Yes | Yes |
| 12 | No | Yes |
| 12 | Yes | Yes |

Content Curated by Pollux M. Rey

Logistic Regression Evaluation Metrics

Confusion Matrix

A table used to evaluate model performance, revealing insights into errors and weaknesses.

| | | Predicted | |
|--------|----------|----------------|----------------|
| | | Positive | Negative |
| Actual | Positive | True Positive | False Negative |
| | Negative | False Positive | True Negative |

Logistic Regression Evaluation Metrics

| Study Hours (x) | Passed? (y) | Predictions (y') |
|--------------------|----------------|---------------------|
| 5 | No | No |
| 1 | No | No |
| 2 | Yes | No |
| 3 | No | No |
| 3 | No | No |
| 4 | No | No |
| 6 | Yes | Yes |
| 7 | Yes | Yes |
| 8 | No | Yes |
| 9 | Yes | Yes |
| 10 | Yes | Yes |
| 11 | Yes | Yes |
| 11 | Yes | Yes |
| 12 | No | Yes |
| 12 | Yes | Yes |

Predicted

| | | Predicted | |
|--------|----------|-----------|----------|
| | | Positive | Negative |
| Actual | Positive | | |
| | Negative | | |

Logistic Regression Evaluation Metrics

| Study Hours (x) | Passed? (y) | Predictions (y') |
|--------------------|----------------|---------------------|
| 5 | No | No |
| 1 | No | No |
| 2 | Yes | No |
| 3 | No | No |
| 3 | No | No |
| 4 | No | No |
| 6 | Yes | Yes |
| 7 | Yes | Yes |
| 8 | No | Yes |
| 9 | Yes | Yes |
| 10 | Yes | Yes |
| 11 | Yes | Yes |
| 11 | Yes | Yes |
| 12 | No | Yes |
| 12 | Yes | Yes |

Predicted

| | | Predicted |
|--------|----------|-----------|
| | | Positive |
| Actual | Positive | 7 |
| | Negative | |

Logistic Regression Evaluation Metrics

| Study Hours (x) | Passed? (y) | Predictions (y') |
|--------------------|----------------|---------------------|
| 5 | No | No |
| 1 | No | No |
| 2 | Yes | No |
| 3 | No | No |
| 3 | No | No |
| 4 | No | No |
| 6 | Yes | Yes |
| 7 | Yes | Yes |
| 8 | No | Yes |
| 9 | Yes | Yes |
| 10 | Yes | Yes |
| 11 | Yes | Yes |
| 11 | Yes | Yes |
| 12 | No | Yes |
| 12 | Yes | Yes |

Predicted

| | | Predicted | |
|--------|----------|-----------|----------|
| | | Positive | Negative |
| Actual | Positive | 7 | 1 |
| | Negative | | |

Logistic Regression Evaluation Metrics

| Study Hours (x) | Passed? (y) | Predictions (y') |
|--------------------|----------------|---------------------|
| 5 | No | No |
| 1 | No | No |
| 2 | Yes | No |
| 3 | No | No |
| 3 | No | No |
| 4 | No | No |
| 6 | Yes | Yes |
| 7 | Yes | Yes |
| 8 | No | Yes |
| 9 | Yes | Yes |
| 10 | Yes | Yes |
| 11 | Yes | Yes |
| 11 | Yes | Yes |
| 12 | No | Yes |
| 12 | Yes | Yes |

Actual

Predicted

| | | Positive | Negative |
|----------|----------|----------|----------|
| Positive | Positive | 7 | 1 |
| | Negative | 2 | |

Logistic Regression Evaluation Metrics

| Study Hours (x) | Passed? (y) | Predictions (y') |
|--------------------|----------------|---------------------|
| 5 | No | No |
| 1 | No | No |
| 2 | Yes | No |
| 3 | No | No |
| 3 | No | No |
| 4 | No | No |
| 6 | Yes | Yes |
| 7 | Yes | Yes |
| 8 | No | Yes |
| 9 | Yes | Yes |
| 10 | Yes | Yes |
| 11 | Yes | Yes |
| 11 | Yes | Yes |
| 12 | No | Yes |
| 12 | Yes | Yes |

Predicted

| | | Predicted | |
|--------|----------|-----------|----------|
| | | Positive | Negative |
| Actual | Positive | 7 | 1 |
| | Negative | 2 | 5 |

Logistic Regression Evaluation Metrics

| | | Predicted | |
|--------|----------|-----------|----------|
| | | Positive | Negative |
| Actual | Positive | 7 | 1 |
| | Negative | 2 | 5 |

Logistic Regression Evaluation Metrics

Accuracy

Total number of correct classifications divided by the total number of cases.

$$\frac{\text{correct classifications}}{\text{total classifications}} = \frac{TP + TN}{TP + TN + FP + FN}$$

<https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>

<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

Logistic Regression Evaluation Metrics

$$\begin{aligned}\text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\ &= \frac{7 + 5}{7 + 5 + 2 + 1} \\ &= \frac{12}{15} \\ &= 0.80\end{aligned}$$

| | | Predicted | |
|--------|----------|-----------|----------|
| | | Positive | Negative |
| Actual | Positive | 7 | 1 |
| | Negative | 2 | 5 |

The model is 80% accurate on the test data.

<https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>

<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

Logistic Regression Evaluation Metrics

Being heavily reliant
on the accuracy metric
can lead to incorrect decisions.

<https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>

<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

Logistic Regression Evaluation Metrics

Consider a model predicting if a patient has an infectious virus with 99.9% accuracy.

If it mistakenly predicts a patient as negative, the virus could spread rapidly.

| | | Predicted | |
|--------|----------|-----------|----------|
| | | Positive | Negative |
| Actual | Positive | 1 | 1 |
| | Negative | 0 | 998 |

<https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>

<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

Logistic Regression Evaluation Metrics

Precision

Total number of true positives divided by
the total number of predicted positives.

$$\frac{\text{correct classified actual positives}}{\text{everything classified as positive}} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

<https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>

<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

Logistic Regression Evaluation Metrics

$$\begin{aligned}\text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ &= \frac{7}{7 + 2} \\ &= \frac{7}{9} \\ &= 0.78\end{aligned}$$

| | | Predicted | |
|--------|----------|-----------|----------|
| | | Positive | Negative |
| Actual | Positive | 7 | 1 |
| | Negative | 2 | 5 |

The model has a precision of 78% on the test data.

<https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>

<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

Logistic Regression Evaluation Metrics

Recall

Total number of true positives divided by
the total number of actual positives.

$$\frac{\text{correct classified actual positives}}{\text{all actual positives}} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

<https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>

<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

Logistic Regression Evaluation Metrics

$$\begin{aligned}\text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\&= \frac{7}{7 + 1} \\&= \frac{7}{8} \\&= 0.88\end{aligned}$$

| | | Predicted | |
|--------|----------|-----------|----------|
| | | Positive | Negative |
| Actual | Positive | 7 | 1 |
| | Negative | 2 | 5 |

The model has a recall of 88% on the test data.

<https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>

<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

Logistic Regression Evaluation Metrics

| Metric | Guidance |
|----------|--|
| Accuracy | <ol style="list-style-type: none">1. Use as a rough indicator of model training progress for balanced datasets.2. Combine with other metrics for performance evaluation.3. Avoid for imbalanced datasets; use alternative metrics. |

| Metric | Guidance | Scenario |
|-----------|---|---|
| Precision | Use when it's very important for positive predictions to be accurate. |  A model misidentified a transaction as fraudulent. |
| Recall | Use when false negatives are more expensive than false positives. |  A model misdiagnosed a patient as healthy. |

<https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>

<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

Logistic Regression Evaluation Metrics

F1 Score

Harmonic mean (a kind of average) of precision and recall.

$$2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} = \frac{2 * \text{TP}}{(2 * \text{TP}) + \text{FP} + \text{FN}}$$

<https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>

<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

Logistic Regression Evaluation Metrics

$$\begin{aligned}\text{F1 Score} &= \frac{2 * \text{TP}}{(2 * \text{TP}) + \text{FP} + \text{FN}} \\ &= \frac{2 * 7}{(2 * 7) + 2 + 1} \\ &= \frac{14}{17} \\ &= 0.82\end{aligned}$$

| | | Predicted | |
|--------|----------|-----------|----------|
| | | Positive | Negative |
| Actual | Positive | 7 | 1 |
| | Negative | 2 | 5 |

The model has an F1 score of 82% on the test data.

<https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>

<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

Logistic Regression Evaluation Metrics



Content Curated by Pollux M. Rev

https://www.youtube.com/watch?v=jr_BcU4QINE

THANK YOU!

Content Curated by Pollux M. Rey