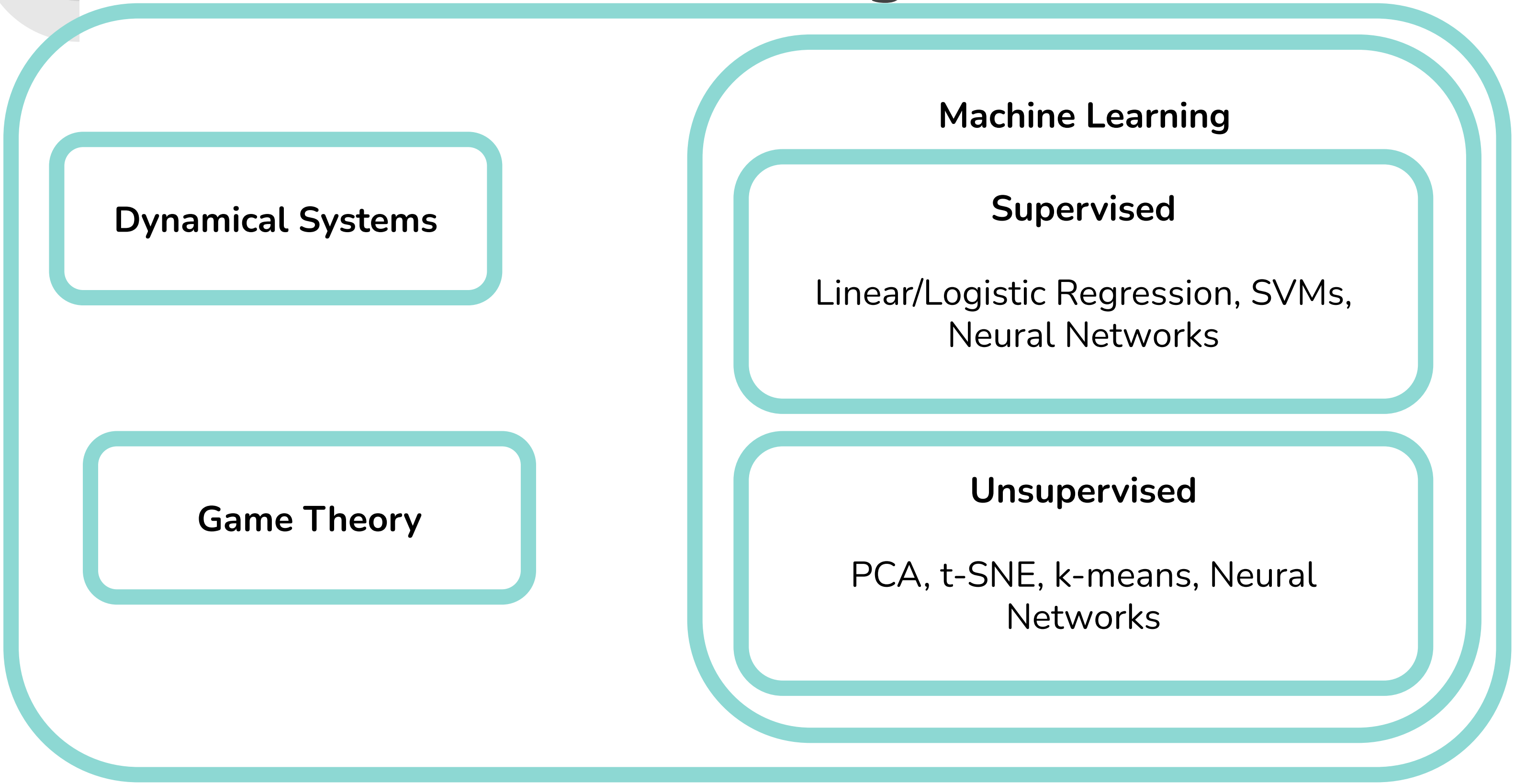
Linear and Logistic Regression

Guillem & Roderic, Summer 2025

Mathematical Modeling



Learning Types

- **Supervised**:
  - Classification (logistic regression, Random Forests, SVM, NNs)
  - Regression (linear regression, Random Forests, NNs)
- **Unsupervised**:
  - Dimensionality Reduction (PCA, t-SNE, UMAP)
  - Clustering (k-means)
- **Other:**
  - Reinforcement Learning (Q-learning, PPO)
  - ...

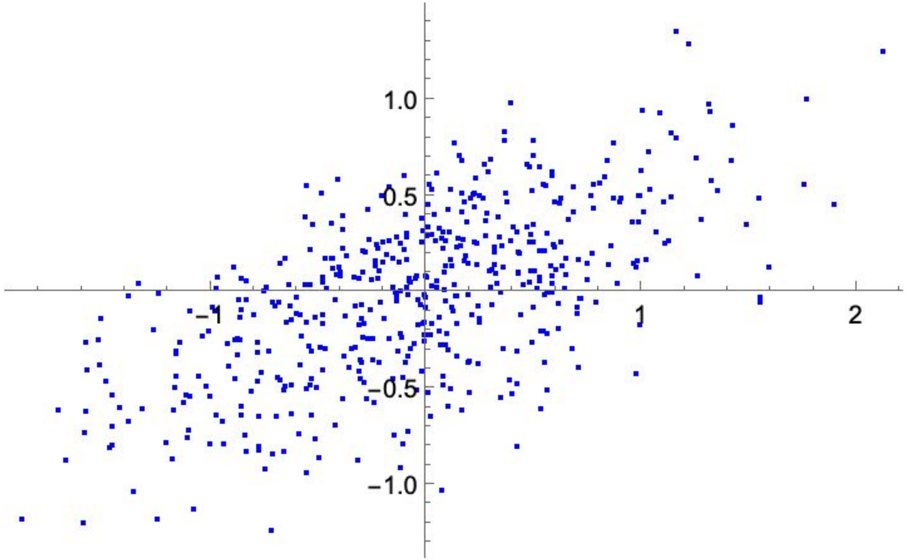
Data Types

- **Numerical**:
  - Price of a home
  - Quality of a Wine
- **Categorical**:
  - Edible and poisonous mushrooms
  - Survival in the Titanic
  - Iris Type
  - Handwritten digit

Linear Regression

Linear Regression

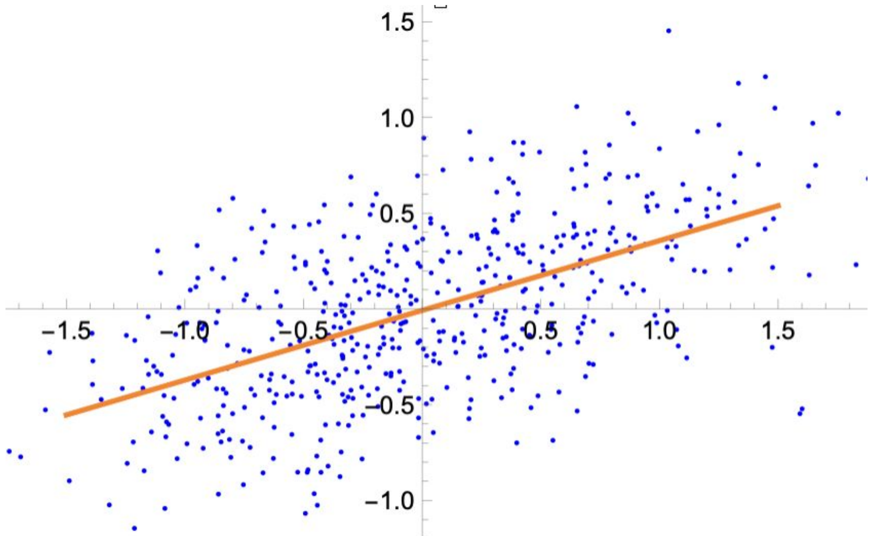
Suppose we have a dataset with two (numerical) variables:



$$\begin{bmatrix} (X_1, Y_1) \\ (X_2, Y_2) \\ (X_3, Y_3) \\ \vdots \\ (X_m, Y_m) \end{bmatrix}$$

Linear Regression

How can we model the linear dependency of **Y** on **X**



Predicted output

Input

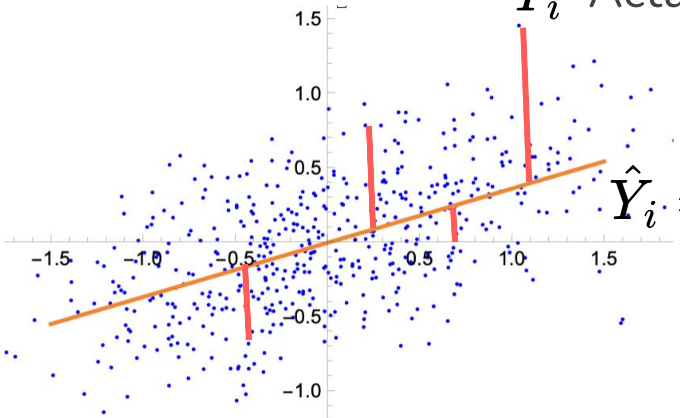
$$Y = mX + b$$

Linear Function

Linear Regression

Minimize **residual sum of squares** (vertical distance): $Y = mX + b$

Y_i Actual value

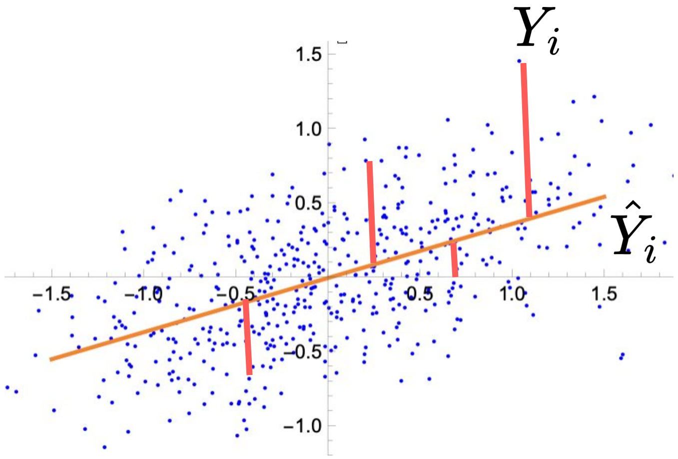


$\hat{Y}_i = mX_i + b$ Predicted value

Linear Regression

Minimize **residual sum of squares** (vertical distance):

$$Y = mX + b$$



$$\hat{Y}_i = mX_i + b$$

Goal: **minimize** the prediction error (residual sum of squares)

$$RSS = \sum_{i=1}^{m} \left(\hat{Y}_i - Y_i\right)^2$$

Example: Boston Homes

Description: "Housing values in **suburbs** of Boston."

CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX
0.11069	0	13.89	tract bounds Charles river	0.55 ppm	5.951	93.8	2.8893	5	276
6.39312	0	18.1	tract does not bound Charles river	0.584 ppm	6.162	97.4	2.206	24	666
0.03578	20	3.33	tract does not bound Charles river	0.4429 ppm	7.82	64.5	4.6947	5	216
0.1146	20	6.96	tract does not bound Charles river	0.464 ppm	6.538	58.7	3.9175	3	223
38.3518	0	18.1	tract does not bound Charles river	0.693 ppm	5.453	100	1.4896	24	666
7.75223	0	18.1	tract does not bound Charles river	0.713 ppm	6.301	83.7	2.7831	24	666
0.01096	55	2.25	tract does not bound Charles river	0.389 ppm	6.453	31.9	7.3073	1	300
0.03705	20	3.33	tract does not bound Charles river	0.4429 ppm	6.968	37.2	5.2447	5	216
0.05515	33	2.18	tract does not bound Charles river	0.472 ppm	7.236	41.1	4.022	7	222
28.6558	0	18.1	tract does not bound Charles river	0.597 ppm	5.155	100	1.5894	24	666

Example: Boston Homes

X = avg. number of rooms

Y = median home value (1000$)

Example: Boston Homes

X = avg. number of rooms

Y = median home value (1000$)

$$\hat{Y} = 9.1X - 34.67$$

What is the meaning of the slope 9.1 in the equation $\hat{Y} = 9.1X - 34.67$?

0

Each additional room increases of price by $9100\$$

0%

On average, each additional room increase the price by $9100\$$

0%

The price of a house with one room is 9100

0%

What is the meaning of the intercept -34.67 in the equation $\hat{Y} = 9.1X - 34.67$?

0

The expected price of a house with 0 rooms is $-34670\$$

0%

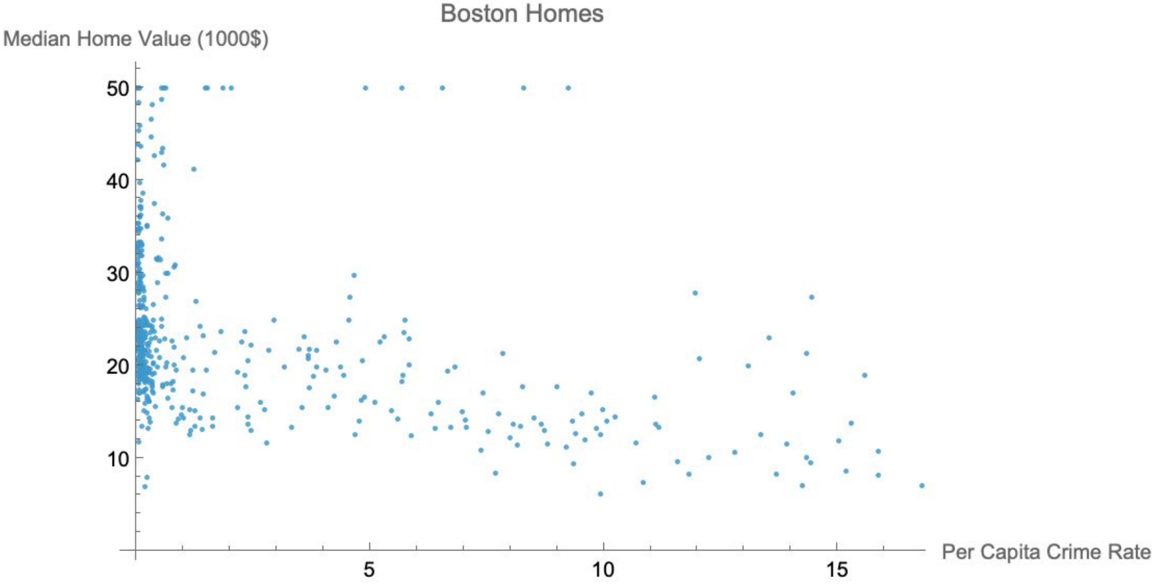
It has no real meaning

0%

If you reduce the number of rooms, the price decreases by $-34670\$$

0%

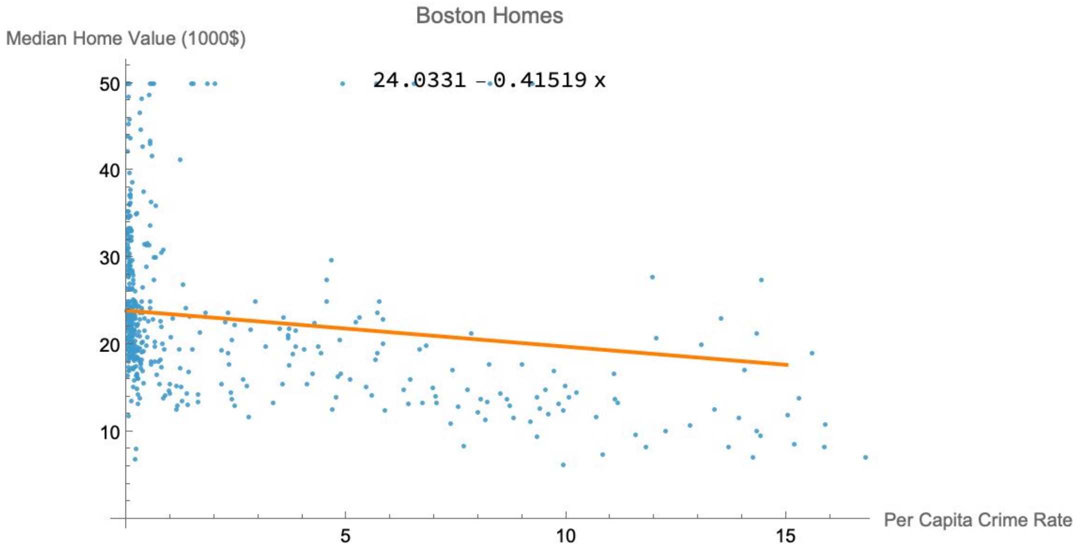
Example: Boston Homes



X = per capita crime rate

Y = median home value (1000$)

Example: Boston Homes



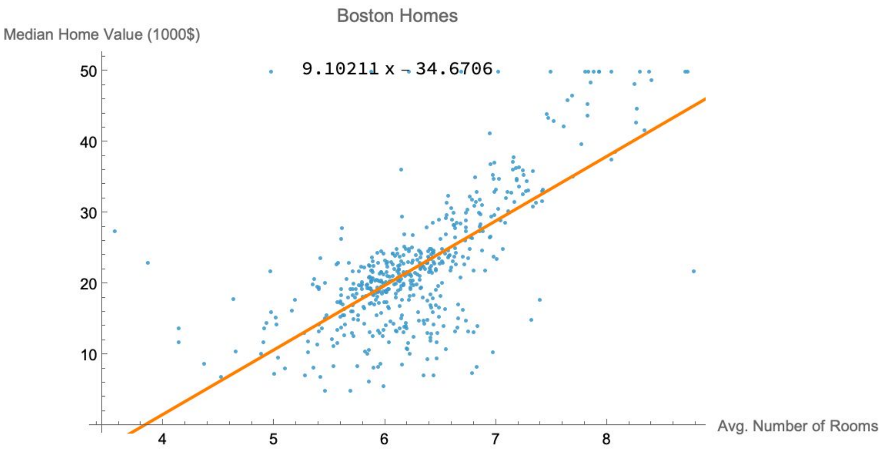
X = per capita crime rate

Y = median home value (1000$)

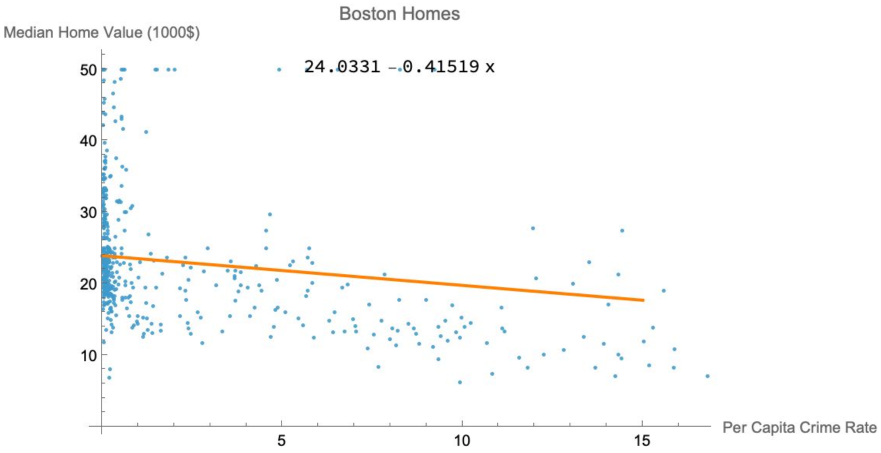
$$\hat{Y} = -0.42X + 24.03$$

Example: Boston Homes

RSS = 148.532



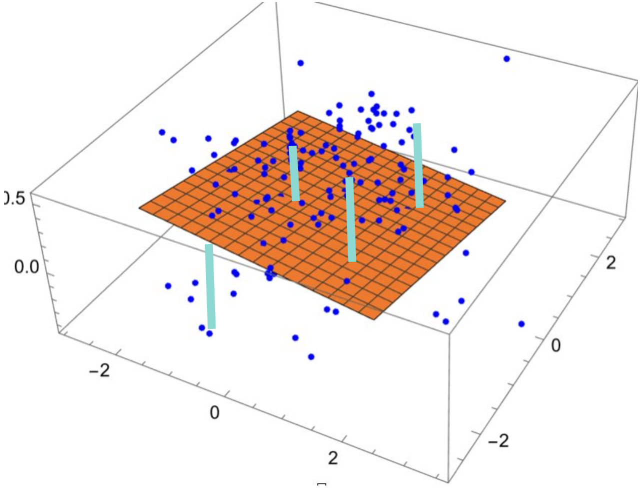
RSS = 190.461



Multidimensional Linear Regression

Suppose we have two (or more) predictor variables

Stil minimize **residual sum of squares** (vertical distances)



$$Y = m_1 X_1 + m_2 X_2 + b$$

If I use two predictors (crime rate and number of rooms) the RSS, when compared to the individual RSSs, will

Increase

Decrease

It depends on the data

If I use two predictors (crime rate and number of rooms) the RSS, when compared to the individual RSSs, will

Increase

0%

Decrease

0%

It depends on the data

0%

If I use two predictors (crime rate and number of rooms) the RSS, when compared to the individual RSSs, will

Increase

0%

Decrease

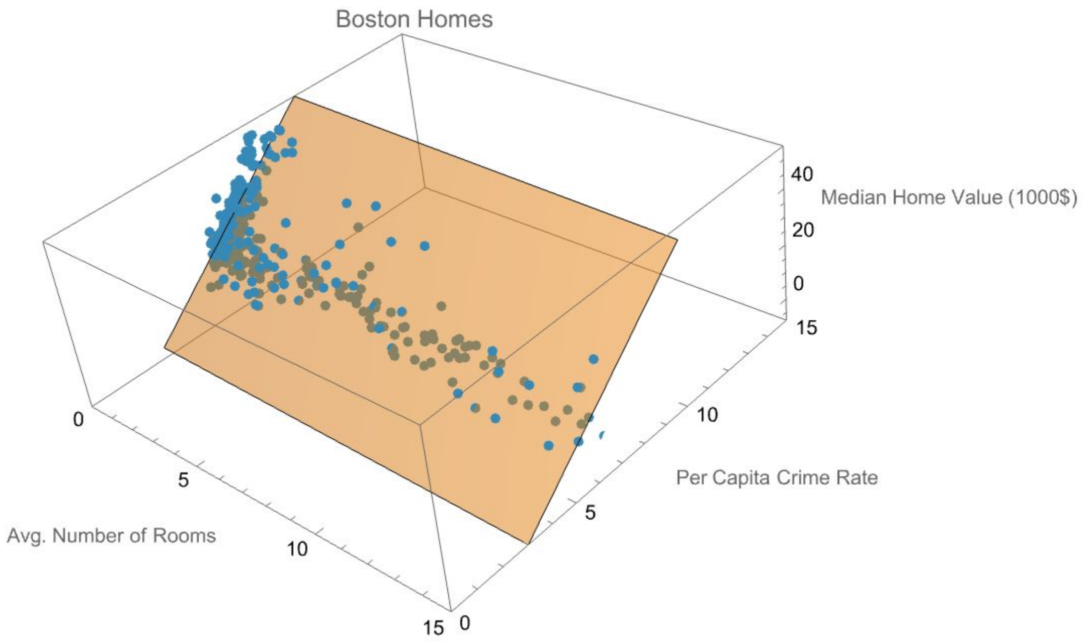
0%

It depends on the data

0%

Example: Boston Homes

$$\hat{Y} = -29.24 - 0.26X_2 + 8.39X_1$$



RSS = 139.878

Example: Boston Homes

Rooms
RSS = 148.532

$$\hat{Y} = 9.1X_1 - 34.67$$

Crime

RSS = 190.461

$$\hat{Y} = -0.42X_2 + 24.03$$

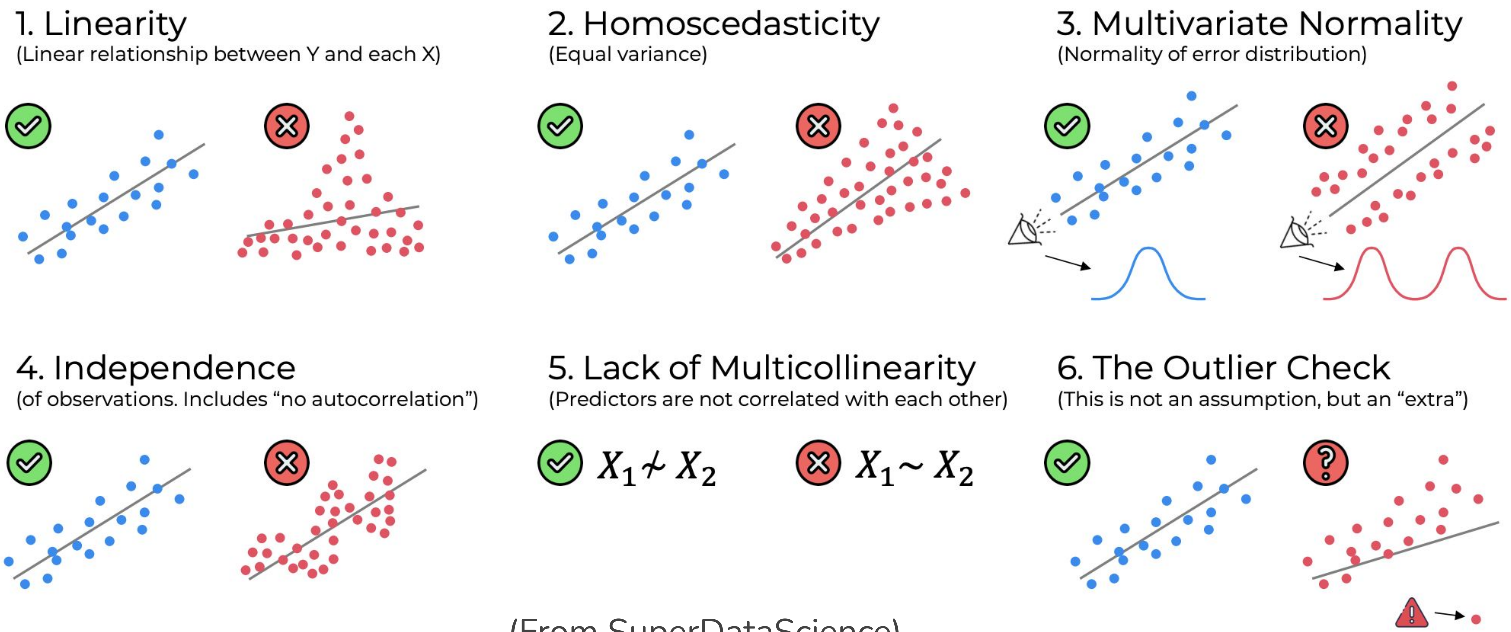
Rooms + Crime

RSS = 139.878

$$\hat{Y} = -29.24 - 0.26X_2 + 8.39X_1$$

Linear Regression: When to Use?

When both the predictor and output are **numerical variables**, and:

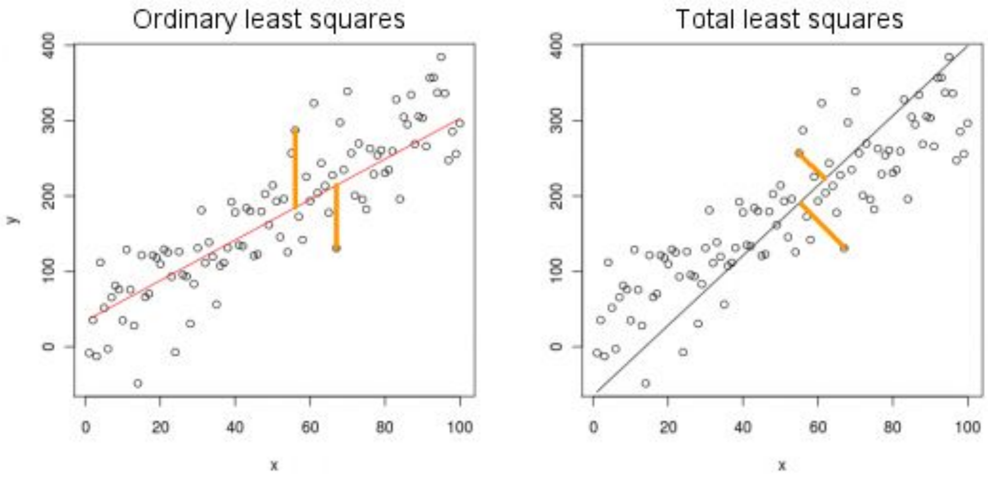


(From SuperDataScience)

Linear Regression vs PCA

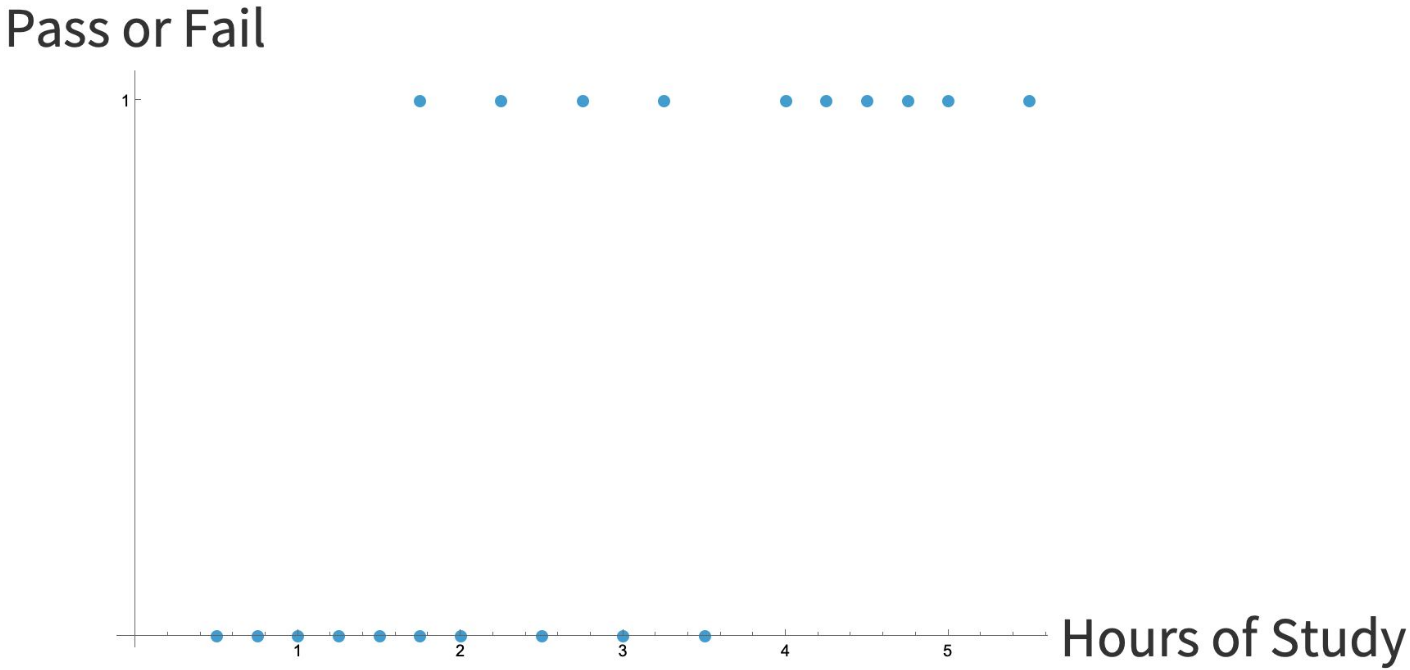
Linear Regression: using X as a predictor, what is the equation that best describes Y

PCA: What linear combination of X and Y (direction/component) is the best predictor of our data?

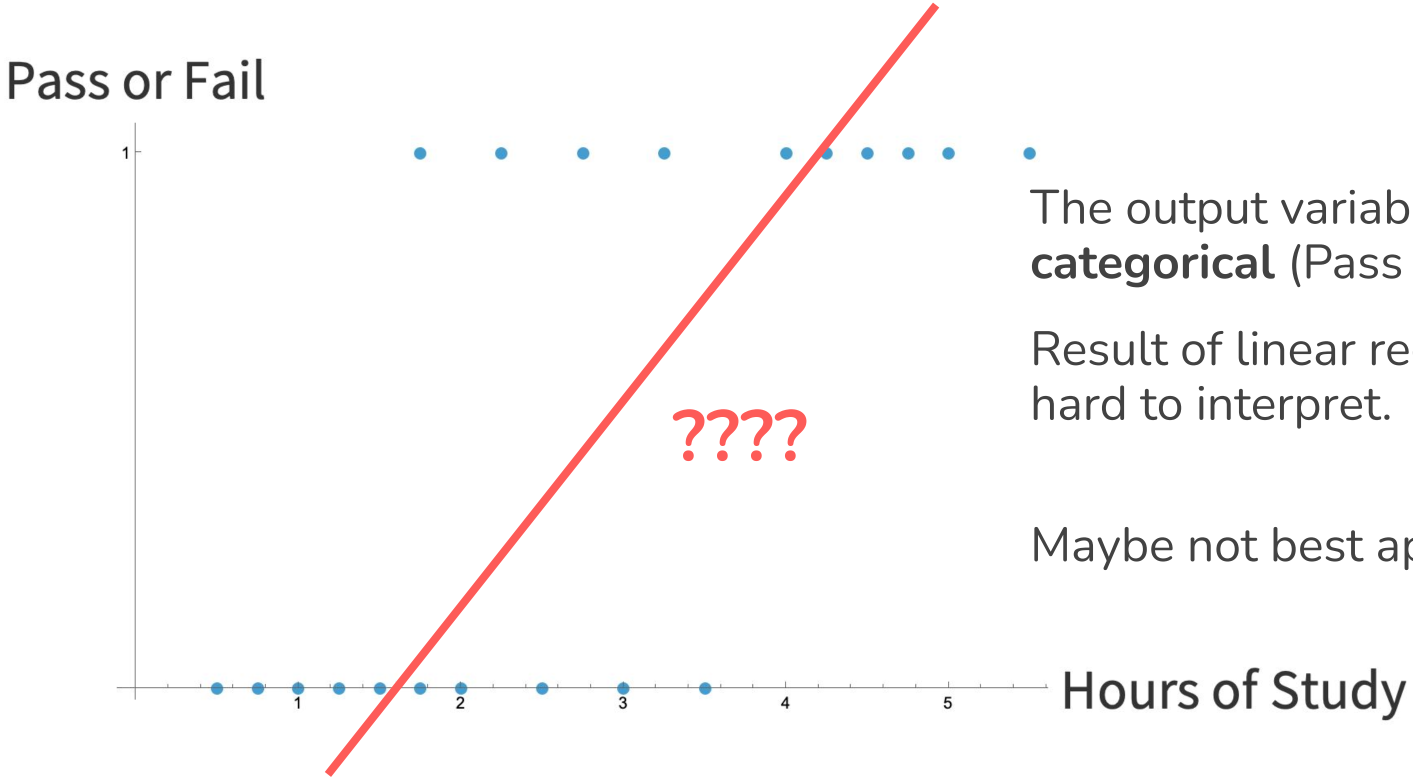


Logistic Regression

Example: Pass/Fail



Example: Pass/Fail



The output variable is **categorical** (Pass or Fail)

Result of linear regression hard to interpret.

Maybe not best approach?

Logistic Regression: Main Idea

Instead of modeling the categorical variable (Pass=1, Fail=0), we model the **probability** of each class:

The probability of passing given you study X hours is $P\left(1 \mid X\right)$

Logistic Regression: The Logit

Probability: P in $(0,1)$

Odds: $\frac{P}{1-P}$ in $(0,\infty)$

Logit (Log Odds): $\text{logit}(P) = \log\left(\frac{P}{1-P}\right)$ in $(-\infty, \infty)$

Logit Regression: $\text{logit}(P) = mX + b$

Solving for P we get the **Sigmoid Function** $P = \frac{1}{1+e^{-mX-b}}$

If $P(1) = 0.75$, what are the $\mathrm{Odds}(1)$?

3

2

1

0.5

If $P(1) = 0.75$, what are the $\mathrm{Odds}(1)$?

3

0%

2

0%

1

0%

0.5

0%

If $P(1) = 0.75$, what are the $\mathrm{Odds}(1)$?

3

0%

2

0%

1

0%

0.5

0%

If $\mathrm{Odds}(1) = 2$, what is $P(1)$?

1/3

2/3

1/2

1/4

If $\mathrm{Odds}(1) = 2$, what is $P(1)$?

1/3

0%

2/3

0%

1/2

0%

1/4

0%

If $\text{Odds}(1) = 2$, what is $P(1)$?

1/3

0%

2/3

0%

1/2

0%

1/4

0%

Logistic Regression

The predicted **P(Y)** value is interpreted as the probability that, given **x**, the **categorical variable Y** belongs to a class 1.

$$\hat{P}(Y = 1 \mid x) = \frac{1}{1 + e^{-(mx+b)}}$$

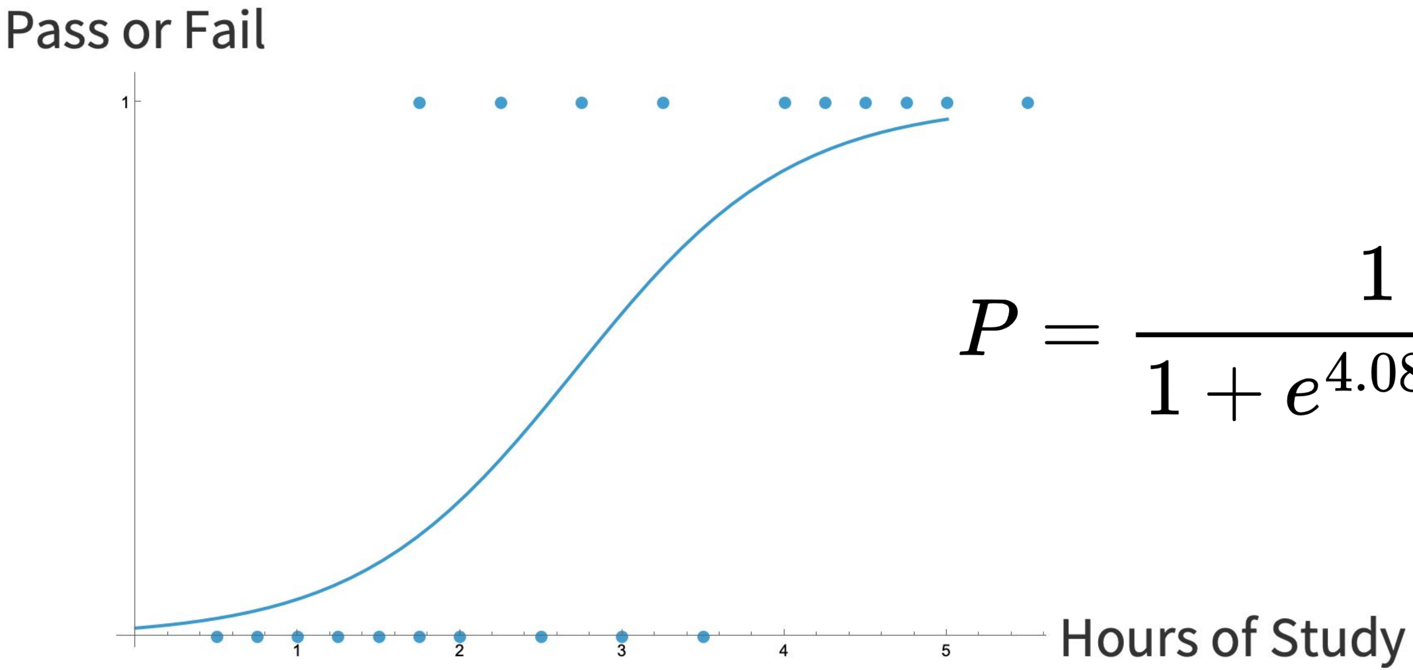
Goal: **minimize** the **Log Loss** $-\sum_{k=1}^{n}\left[Y_k \ln\left(\hat{P}_k\right) + (1 - Y_k)\ln\left(1 - \hat{P}_k\right)\right]$

Logistic Regression: The Quantities

Logit, Odds, And Probability Table

	Probability (p)	Odds (p / (1 - p))	Logit (log(p / (1 - p)))
1	0.01	0.0101	-4.5951
2	0.1	0.1111	-2.1972
3	0.25	0.3333	-1.0986
4	0.5	1.0	0.0
5	0.75	3.0	1.0986
6	0.9	9.0	2.1972
7	0.99	99.0	4.5951

Example: Pass/Fail



$$P = \frac{1}{1 + e^{4.08 - 1.50x}}$$

Consider the logistic model $\frac{1}{1+e^{4.08-1.5x}}$. If you study 4 hours, what is your probability of passing?

78%	0%
87%	0%
91%	0%

Consider the logistic model $\frac{1}{1+e^{4.08-1.5x}}$. The number of hours you should study so that the probability of passing is greater than the probability of failing is:

4.08

0%

-1.5

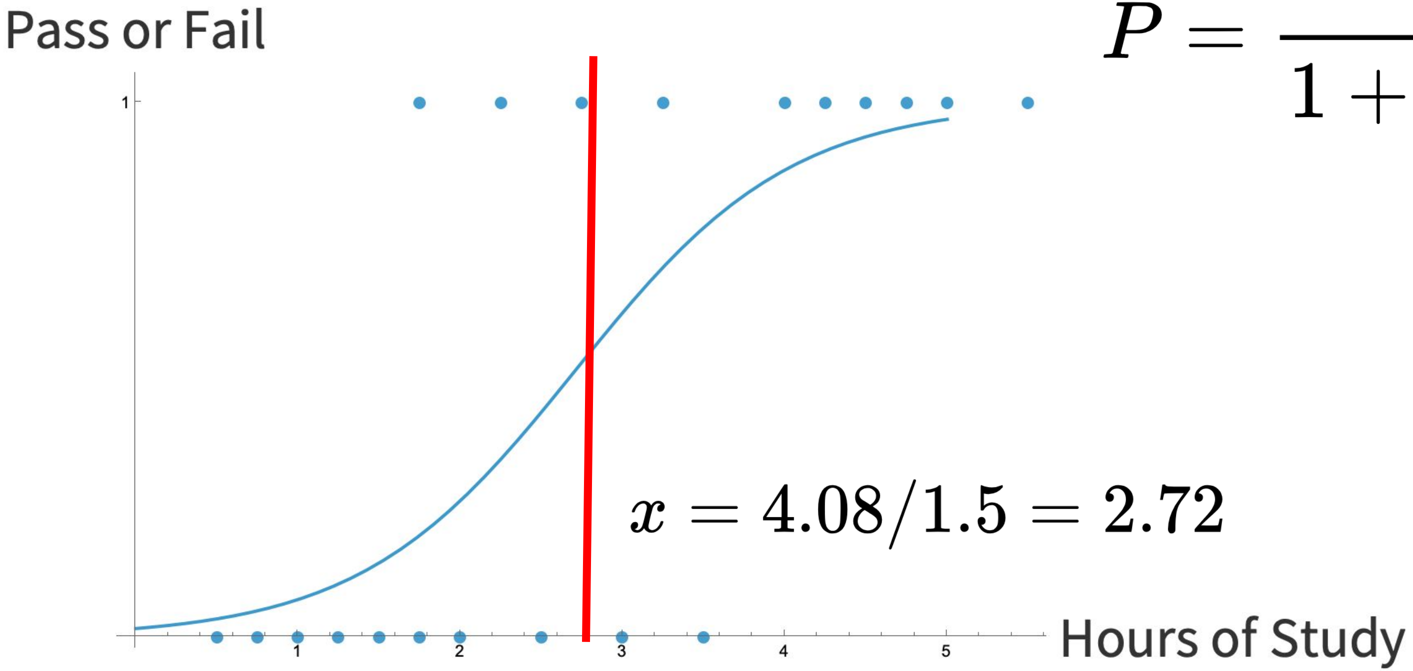
0%

2.72

0%

Example: Pass/Fail

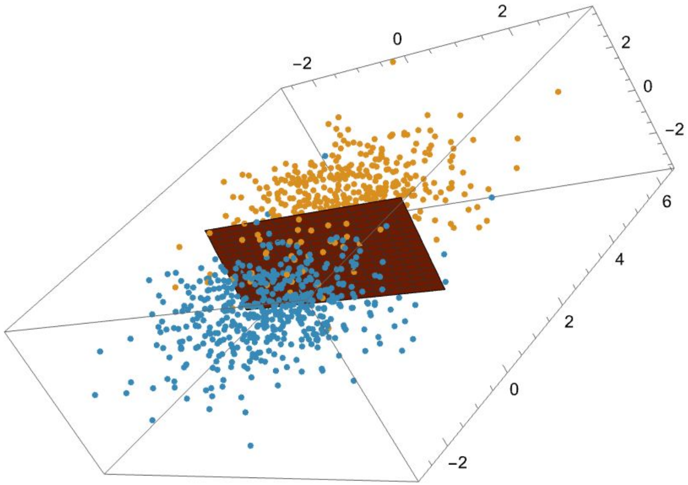
$$P = \frac{1}{1 + e^{4.08 - 1.50x}}$$

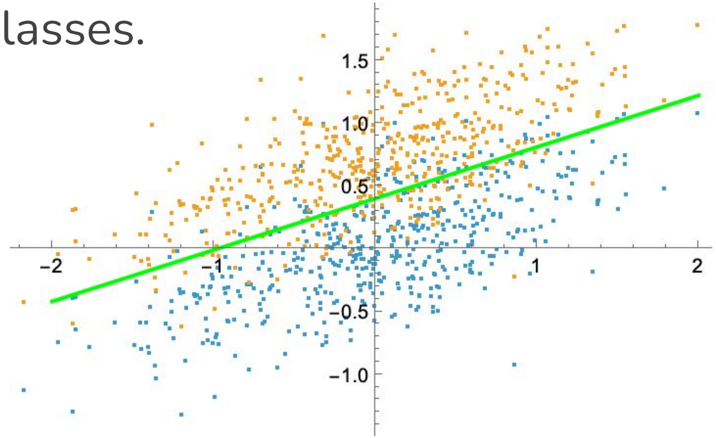


$$x = 4.08/1.5 = 2.72$$

Logistic Regression: Visually

Finds the line/plan that separates two classes.

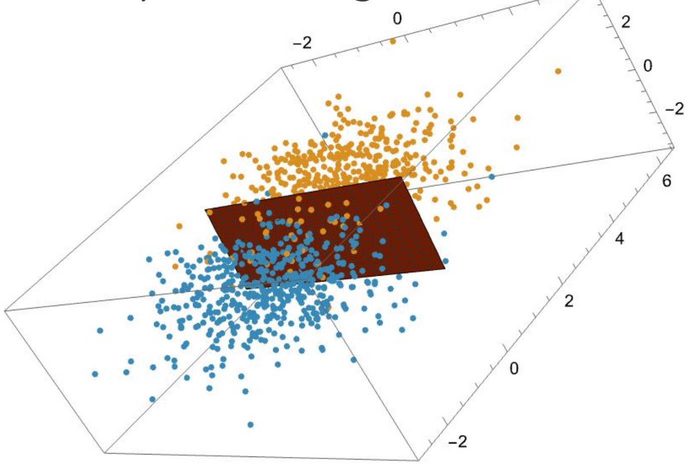


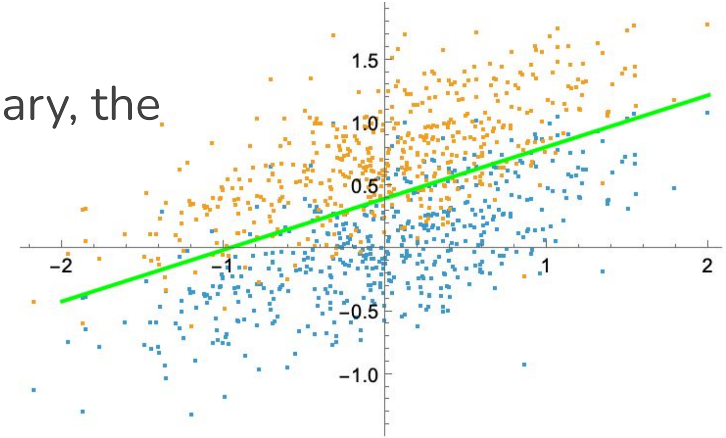


Logistic Regression: Visually

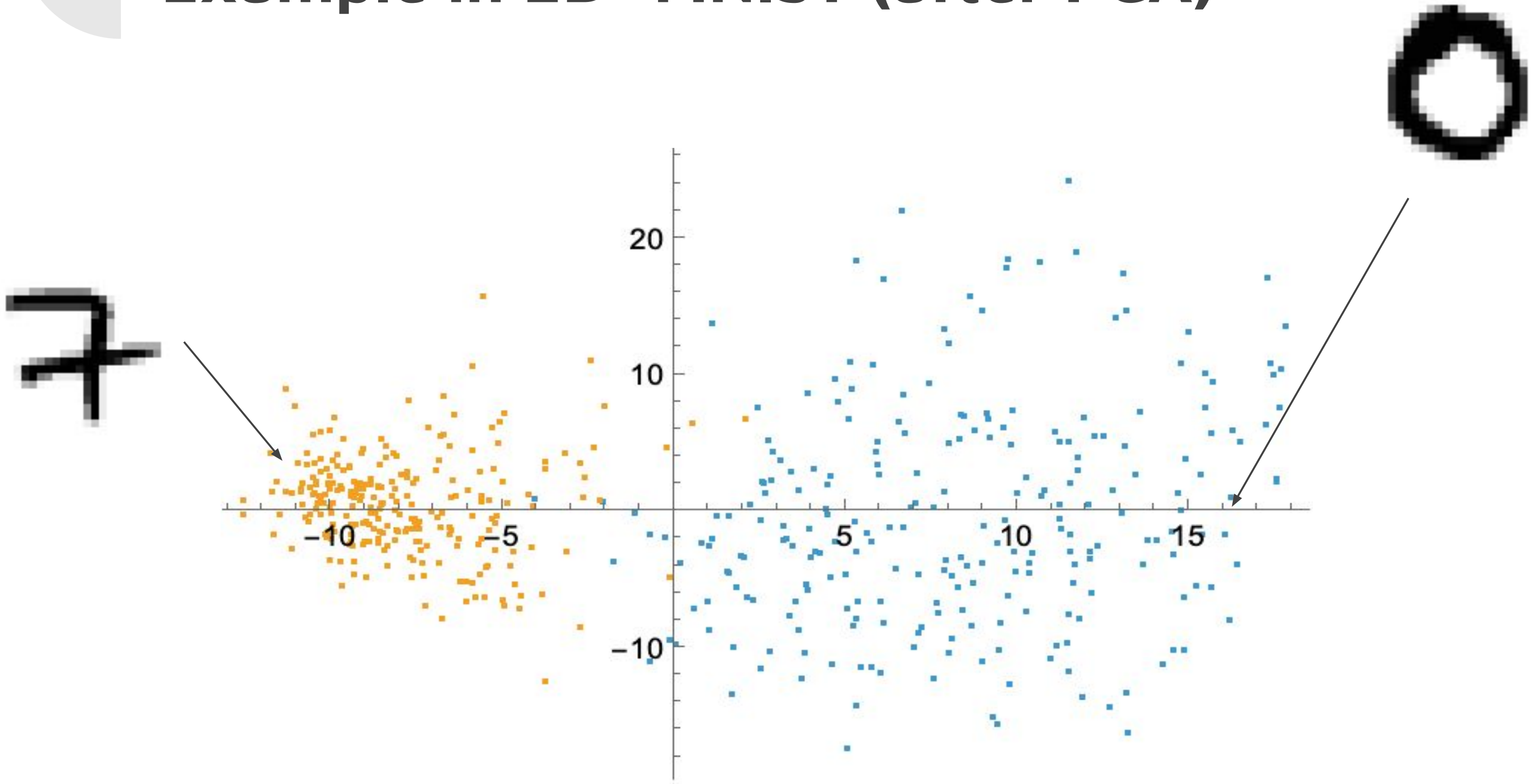
Tells us the likelihood of a point belonging to a class.

The farther the point is from the boundary, the more likely to belong to a class.

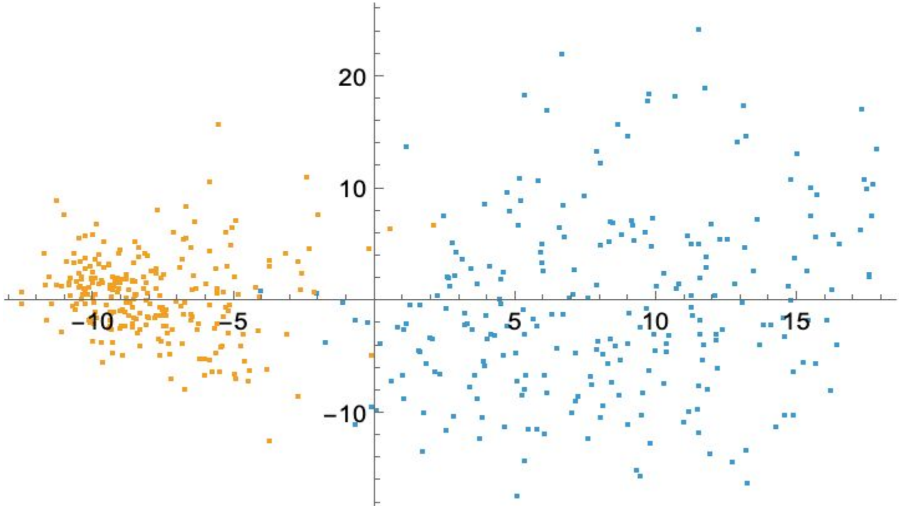


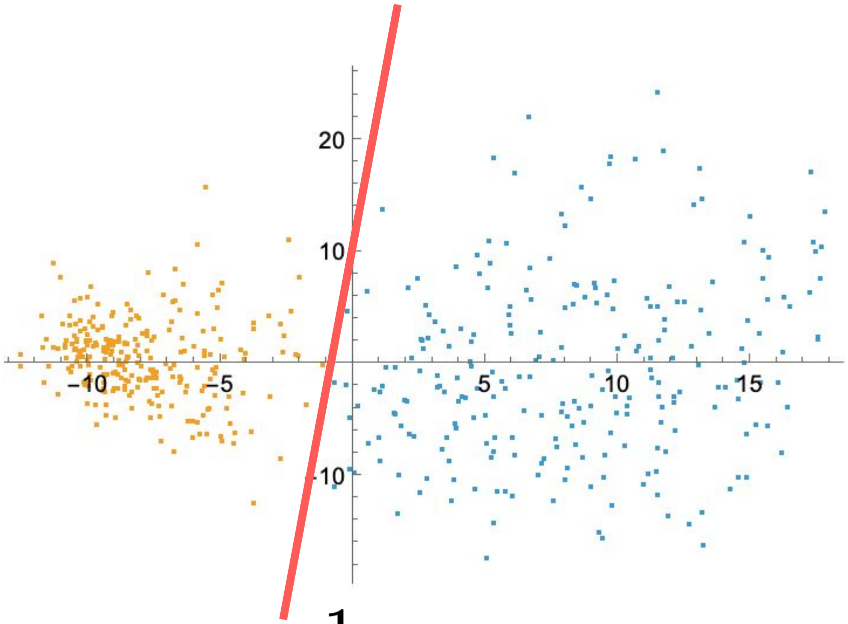


Example in 2D: MNIST (after PCA)



Example in 2D: MNIST (after PCA)

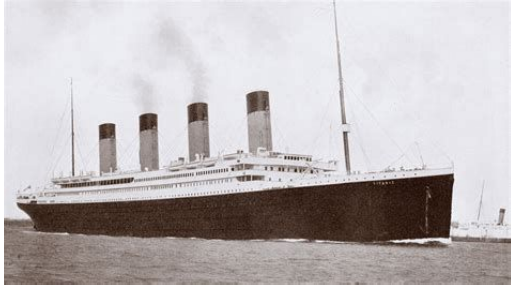




$$\frac{1}{1+e^{1.18434x-0.17231y+1.13553}}$$

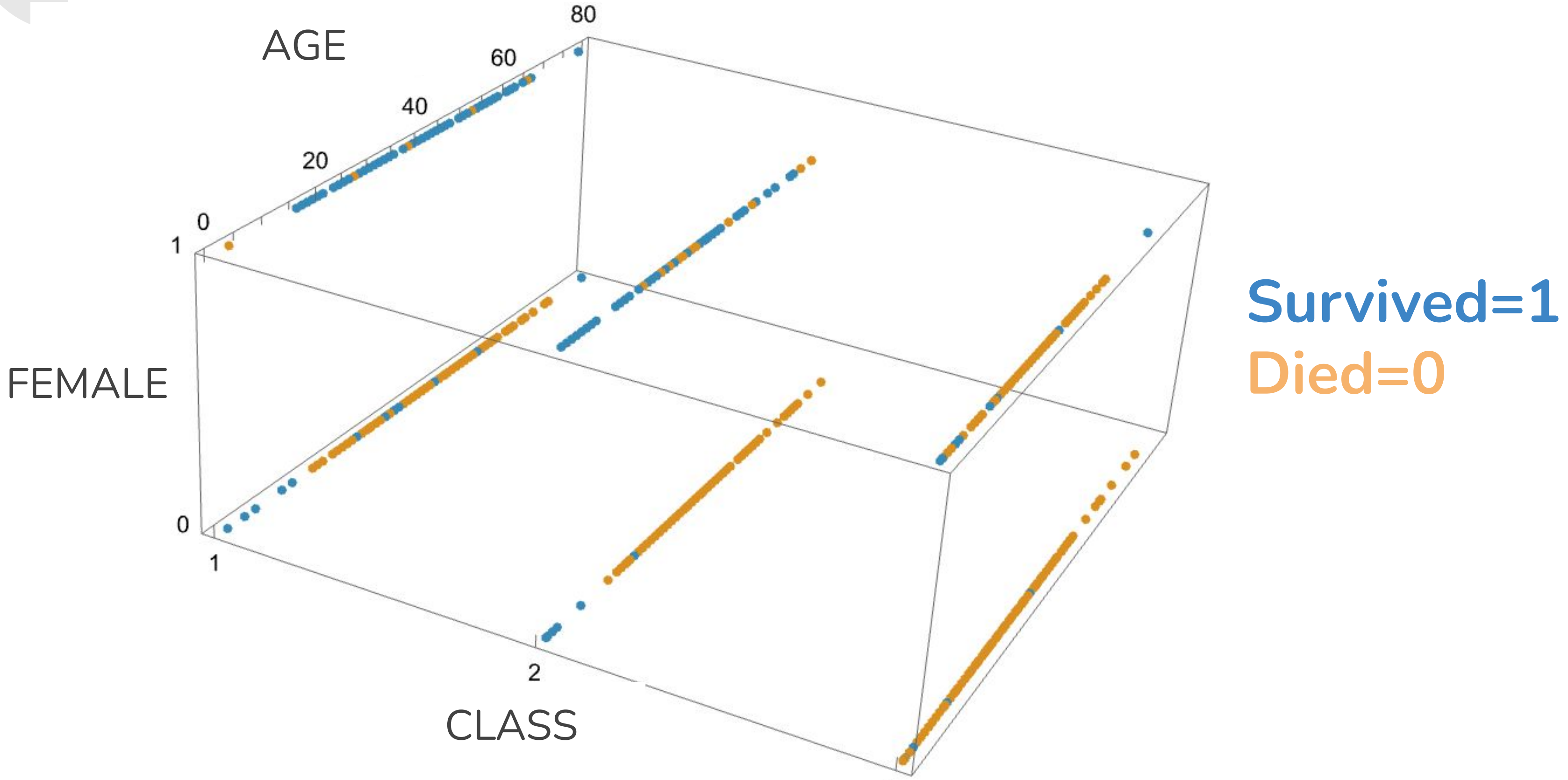
Example in 3D: The Titanic

“Classify whether a passenger on board the maiden voyage of the Titanic in 1912 survived given their age, sex and class.”

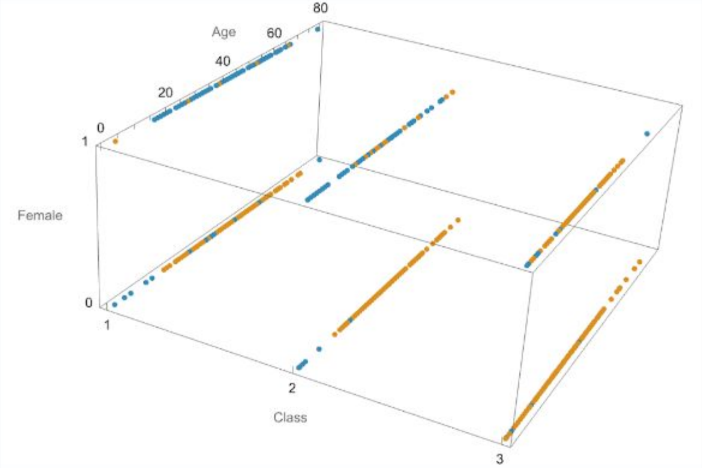


Class	Age	Sex	SurvivalStatus
3rd	12. yr	male	survived
3rd	29. yr	male	died
2nd	28. yr	female	survived
1st	16. yr	female	survived
3rd	—	male	died
3rd	20. yr	male	died
3rd	43. yr	male	died
3rd	18. yr	male	died
3rd	28.5 yr	male	died
1st	—	male	died

Example in 3D: The Titanic



What has higher probability of surviving?



Old male in class 1

0%

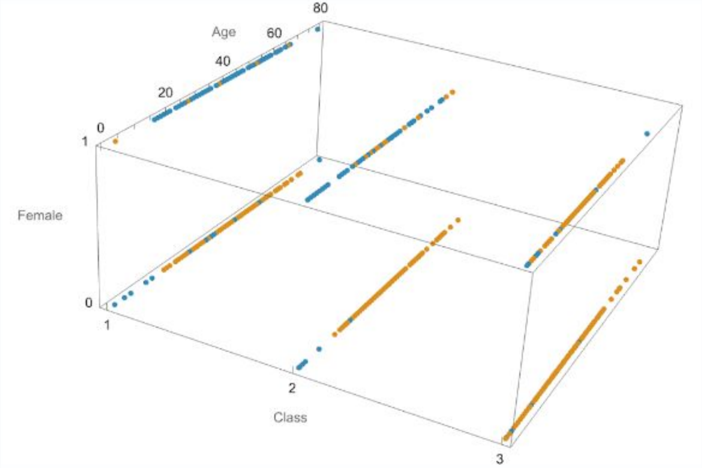
Young female in class 2

0%

Hard to tell

0%

What has higher probability of surviving?



Young male in class 1

0%

Old female in class 3

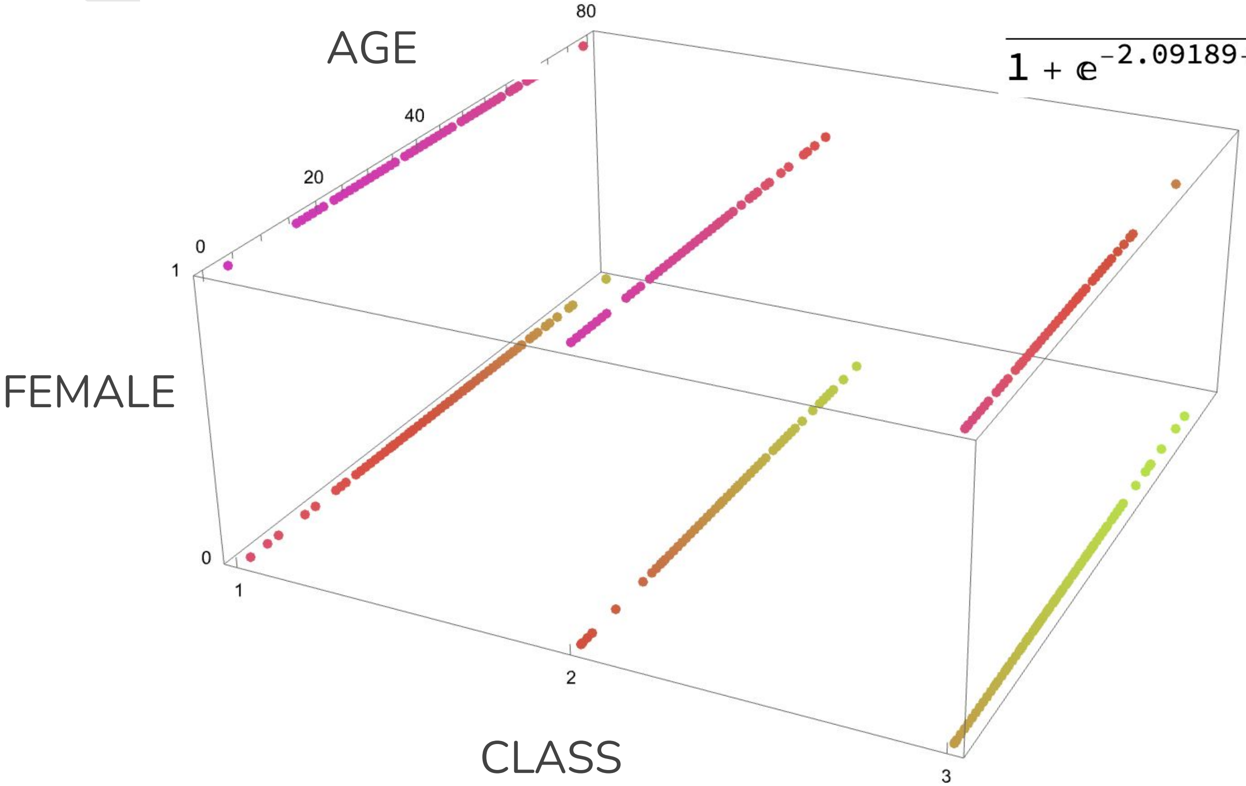
0%

Hard to tell

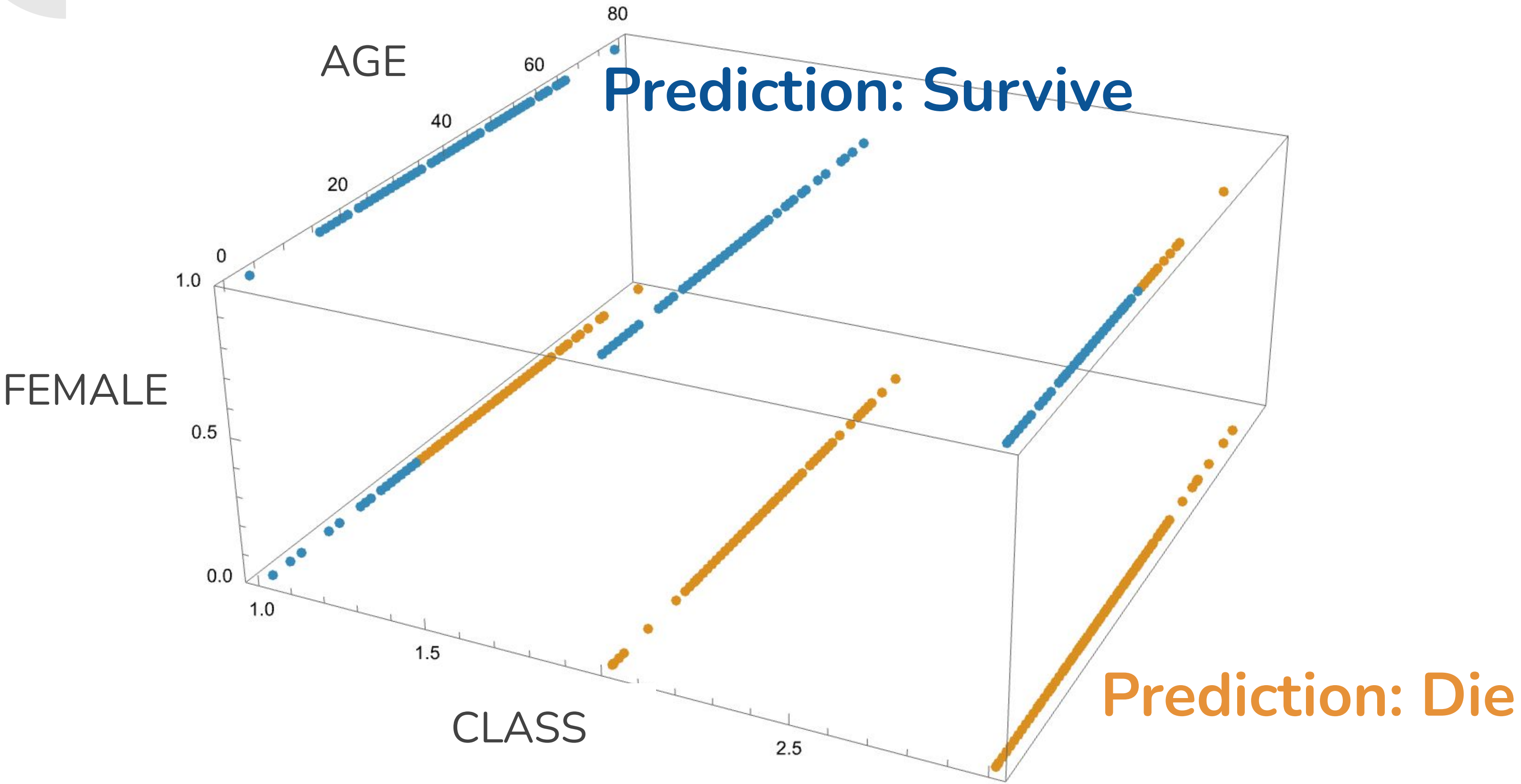
0%

Example in 3D: The Titanic

$$\frac{1}{1 + e^{-2.09189+1.13324\,x+0.033885\,y-2.49738\,z}}$$



Example in 3D: The Titanic



Given this logistic model $\frac{1}{1+e^{-2+x+0.05y-2.5z}}$, where x is class, y is age and z is gender, what is the probability that a 33 year old man in class 2 survived?

22%

35%

52%

Given this logistic model $\frac{1}{1+e^{-2+x+0.05y-2.5z}}$, where x is class, y is age and z is gender, what is the probability that a 33 year old man in class 2 survived?

22%	0%
35%	0%
52%	0%

Given this logistic model $\frac{1}{1+e^{-2+x+0.05y-2.5z}}$, where x is class, y is age and z is gender, what is the probability that a 33 year old man in class 2 survived?

22%	0%
35%	0%
52%	0%

Linear vs Logistic Regression Summary

Linear Regression:

- Purpose:
    - Establish potential relationships between input/output variables
    - Make predictions for newly observed data
    - Best for
        1. Numerical predictor
        2. Numerical output

Logistic Regression:

- Purpose:
    - Estimate the probability that an input belongs to a particular class
    - Classify new data points based on a threshold
    - Best for
        1. Numerical Predictor
        2. Categorical Output