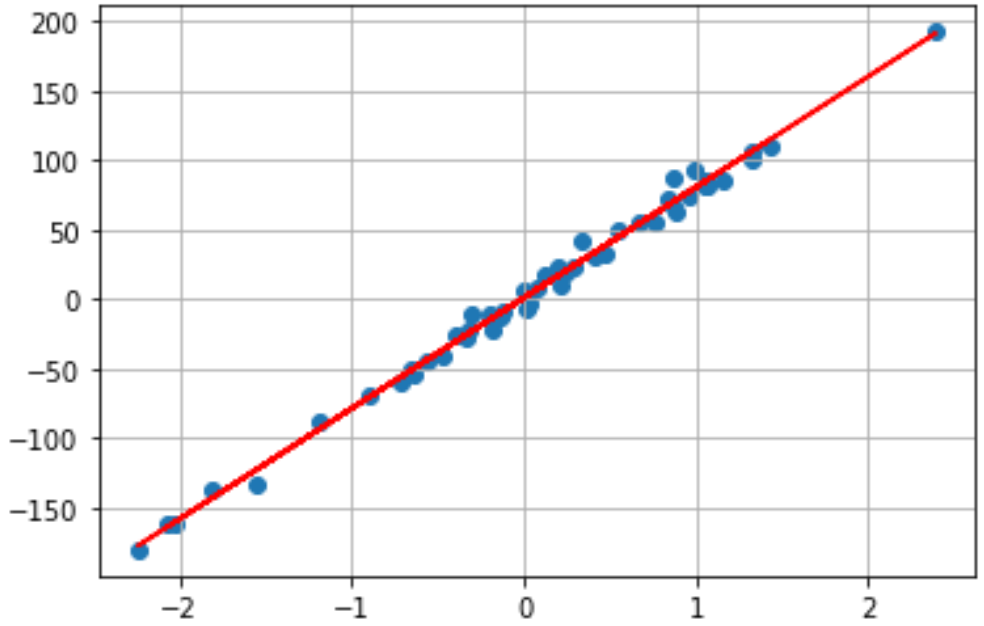
Artificial Intelligence I: Introduction to Data Science and Machine Learning

Assoc. Prof. Dr. Taner Arsan
H. Fuat Alsan, PhD(c)
Sena Kılınç, PhD(c)

Univariate Linear Regression

- $y = w_0 + w_1 x$
- Single input, single output basic model
- x is input (feature)
- y is output (prediction)
- Learnable parameters: w_0, w_1
- w_0 bias (intercept)
- w_1 weight (slope)



Multivariate Linear Regression

- $y = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + \ldots$
- Multiple input, single output model
- $x_1, x_2, x_3, \ldots$ are inputs (features)
- y is output (prediction)
- Learnable parameters: $w_0, w_1, w_2, w_3, \ldots$
- w_0 bias
- w_1, w_2, w_3 weights
- Useful for tabular data

Water (x_1)	pH (x_2)	Sunlight (x_3)	Target (y)
2.22	5.62	128.02	155.0
2.45	5.66	102.55	156.0

Polynomial Regression

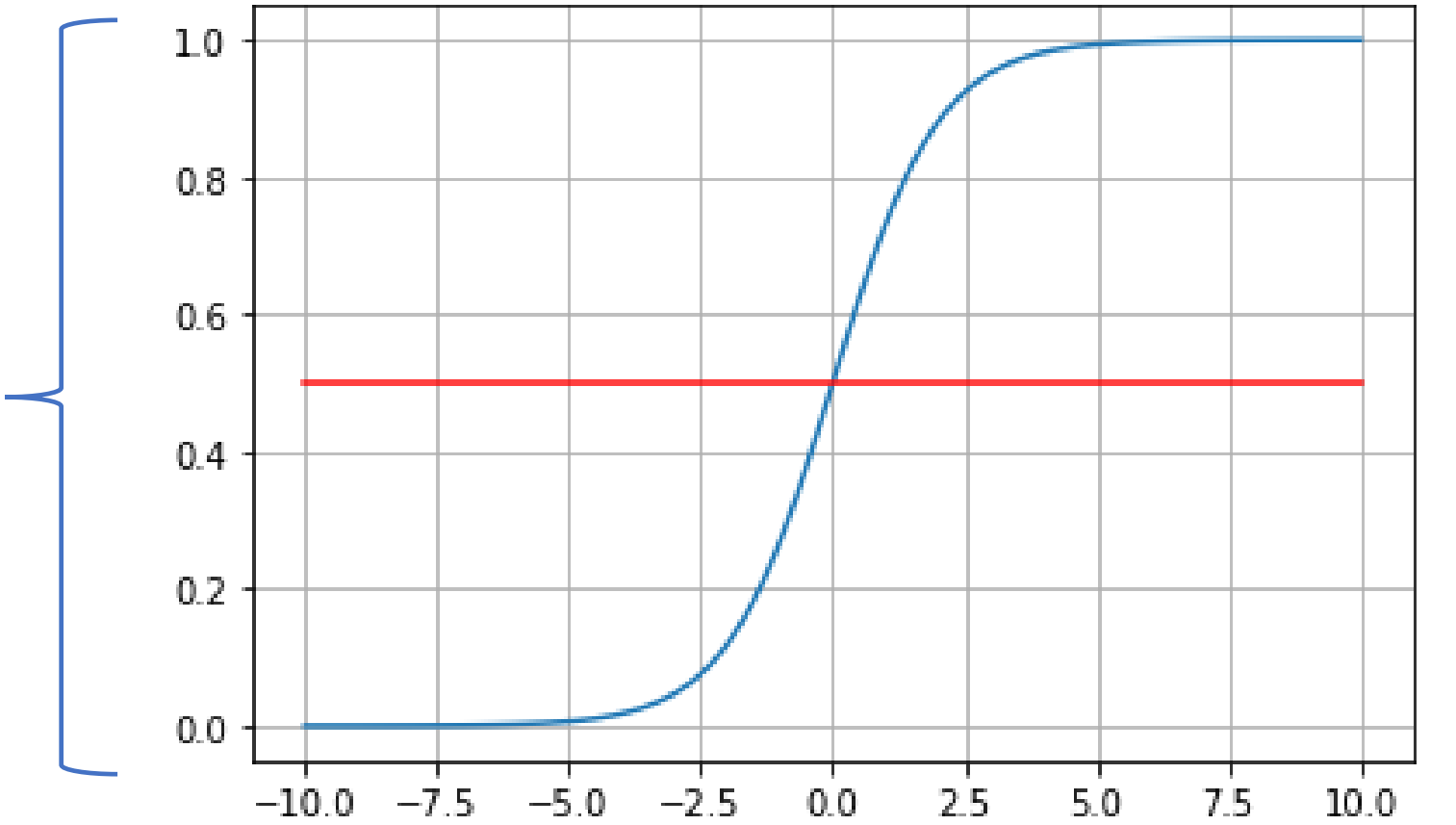
- Polynomial Regression can be implemented with multivariate linear regression with polynomial features
- Polynomial features generation:
  - $x_1, x_2 \rightarrow 1, x_1, x_{2,} x_1^2, x_1 x_2, x_2^2$
- $y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_1 x_2 + w_5 x_2^2$

Logistic Regression

- $y = \sigma(w_0 + w_1 x)$
- Linear regression with sigmoid activation function $\sigma = \frac{1}{1+e^{-x}}$
- Sigmoid will compress values into range [0.0, 1.0]
- Used for classification
- Decision boundary is used to decide which class

Logistic Regression

Sigmoid gives outputs between [0.0, 1.0]



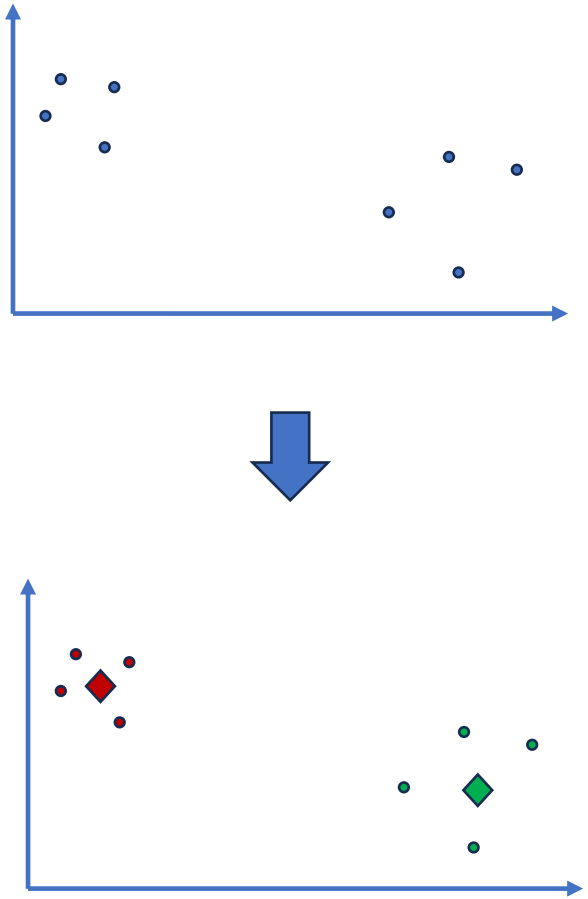
$y > 0.5 \rightarrow Class\ 1$

$Decision\ Boundary\ at\ 0.5$

$y \leq 0.5 \rightarrow Class\ 0$

Clustering with K-Means

- Grouping similar data points together
- Uses a similarity metric
  - (ex: Euclidean distance)
- **Unsupervised** -> Can work with unlabeled data
- Each cluster has a centroid



Common Metrics

	Regression	Classification	Clustering
Training (Loss)	Mean Squarred Error (MSE)	Cross-Entropy	Euclidean Distance
Performance (Score)	R^2	Accuracy, F-1	Silhouette Score