COMP4388: **MACHINE LEARNING**

Feature Scaling

Dr. Radi Jarrar
Department of Computer Science
Birzeit University

Feature Scaling

- Quantitative data is a measure of something: salaries, number of students, number of items, ...
- In machine learning applications, datasets are usually of multiple variables of different scales
- Feature scaling ensures that all features are on the same scale (so that they contribute equally to the distance formula)

Feature Scaling

- This is important for learner that uses distance measure between data points
- Gradient Descent converges much faster with scaled features

Feature Scaling (2)

- The aim is to get features into the range $-1 \leq x_i \leq 1$
- Scaling features means to make all features on the same scale. E.g., if one feature is on the scale 1-10 and another 1,000-10,000; then scaling the features will result in having the second feature on the scale of 1-10 as well

Example

Price(Y)	Area (m^2) x_1	Dis. to CC x_2	Nr. of Roads x_3
40000	600	3000	2
50000	650	1500	2
60000	800	2500	3
100000	1000	100	2
35000	600	5000	1

- It can be well-noticed that the scale of x_2 is much higher than x_3
- Scaling these features is ultimately required

Min-Max scaling

- Min-Max scaling transforms the features so they become in a specific range (i.e., [0, 1])

$$X_{new} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

- It indicates how far (from 0 to 100 percent) the original value fell along between the original min and max values
- The min value becomes zero, the max becomes 1, and all other features become between them. Example?
- Issues: doesn't handle outliers very well

Min-Max scaling (2)

- Issues: doesn't handle outliers very well
- E.g., if you have 99 values between 0 & 50, and one value = 100, then 99 will be transformed to a value between 0 – 0.5
- Accordingly, most of the transformed data is in the range 0.0 – 0.5

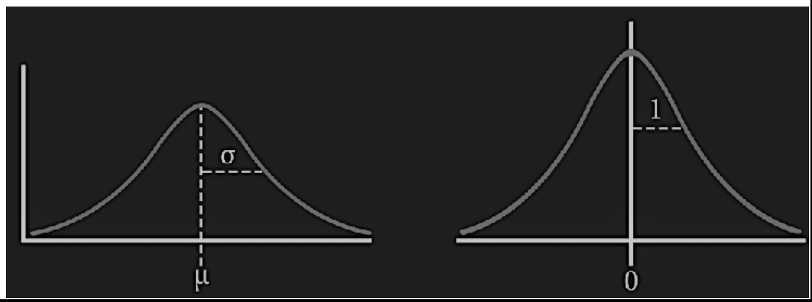
Example – Min-Max

Price(Y)	Area (m^2) x_1	Dis. to CC x_2	Nr. of Roads x_3
40000	0	0.591837	0.5
50000	0.125	0.285714	0.5
60000	0.5	0.4897961	
100000	1	0	0.5
35000	0	1	0

Standardisation (Z-score normalisation)

- Normalises the data using z-score
- Transforms the data such that the resulting distribution will have a mean of zero and standard deviation of 1
- $x_{new} = \frac{x_i - \bar{x}}{s}$

where $\bar{x}$ is the mean and s is the standard deviation



Standardisation (2)

- The transformed feature represents the number of standard deviations the original value is away from the features mean value (i.e., z-score in statistics)
- What if the original feature equal zero?
- This method normalises data and avoids outliers issue
- A problem with z-score: the transformed features are not on the same scale unlike min-max

Standardisation (3)

- Z-score doesn't change the shape of data (normally distributed always?
- Used in PCA, Logistic Regression, SVM, and ANN

Example – z-score normalisation

Price(Y)	Area (m^2) x_1	Dis. to CC x_2	Nr. of Roads x_3
40000	-0.325	0.118367	0
50000	-0.2	-0.18776	0
60000	0.175	0.016327	0.5
100000	0.675	-0.47347	0
35000	-0.325	0.526531	-0.5

