# Steps for Principal Component Analysis

Analytics Vidhya
Learn everything about analytics

# Steps for Principal Component Analysis

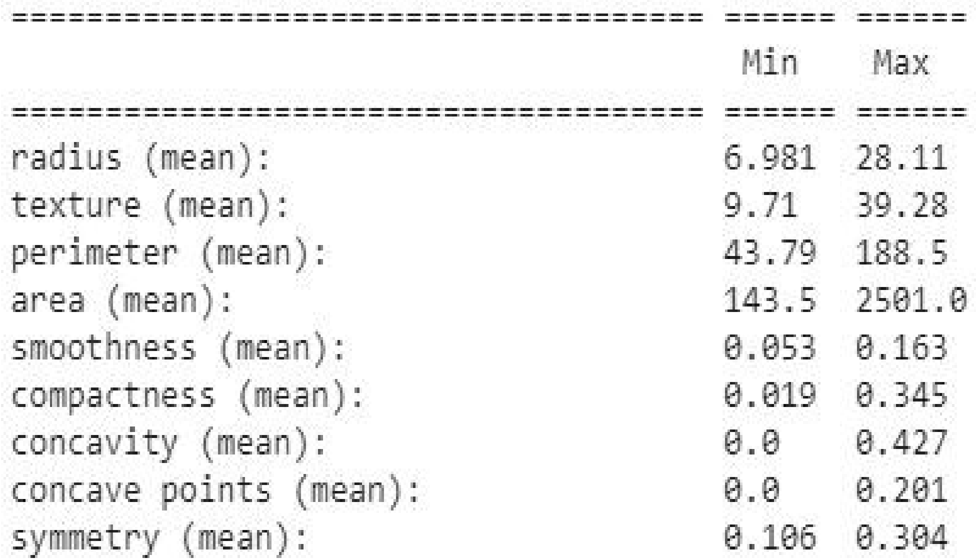
- Standardisation
- Finding the Covariance Matrix
- Finding the directions of maximum variance - Eigenvectors and Eigenvalues
- Selecting the Principal Components

# Standardisation

# Why is standardisation required?

- Different features have different ranges and scales

Consider the features in the famous breast cancer dataset

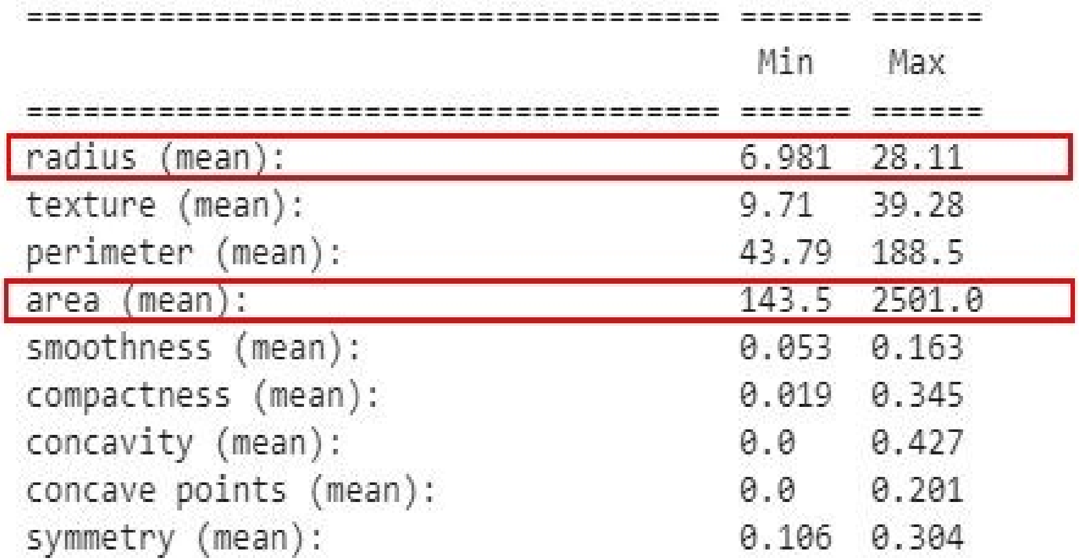


```
===================================== ====== ======
                                        Min    Max
===================================== ====== ======
radius (mean):                        6.981  28.11
texture (mean):                       9.71   39.28
perimeter (mean):                     43.79  188.5
area (mean):                          143.5  2501.0
smoothness (mean):                    0.053  0.163
compactness (mean):                   0.019  0.345
concavity (mean):                     0.0    0.427
concave points (mean):                0.0    0.201
symmetry (mean):                      0.106  0.304
```

# Why is standardisation required?

- Different features have different ranges and scales

Consider the features in the famous breast cancer dataset

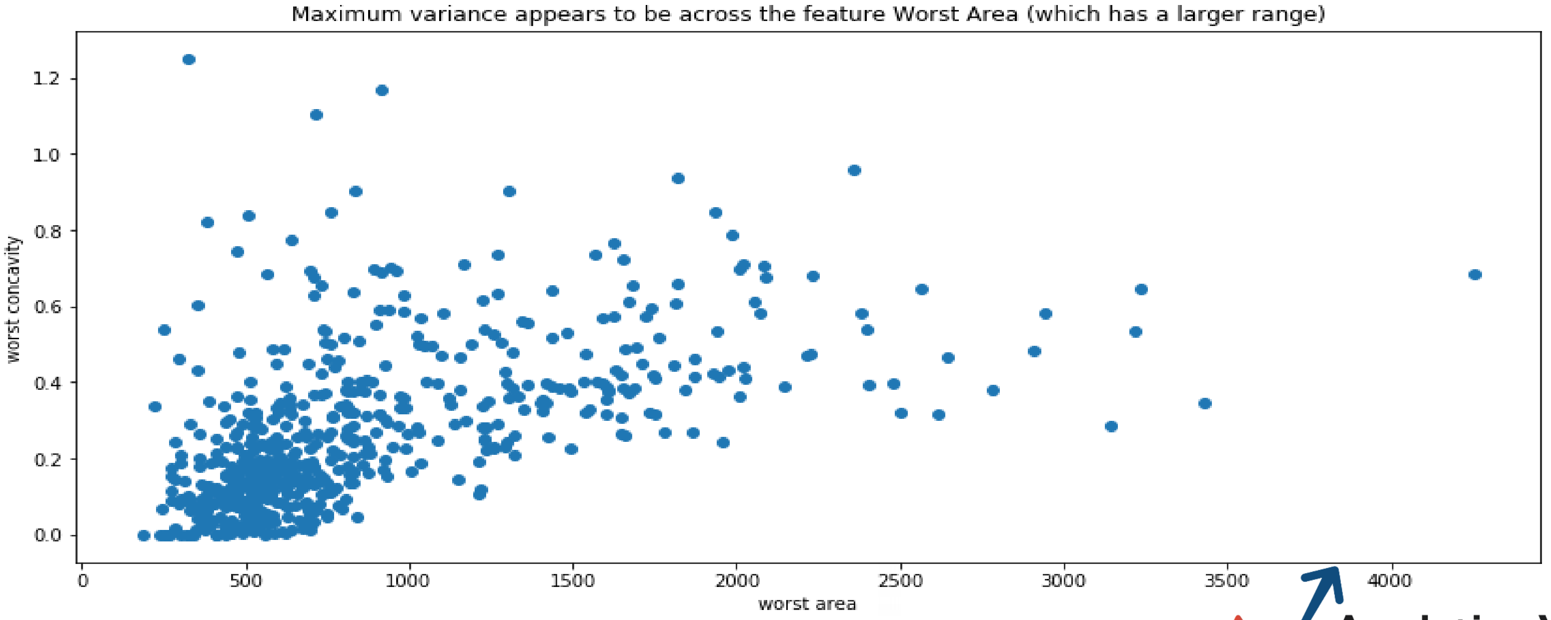


```
===================================== ====== ======
                                       Min    Max
===================================== ====== ======
radius (mean):                         6.981  28.11
texture (mean):                        9.71   39.28
perimeter (mean):                      43.79  188.5
area (mean):                           143.5  2501.0
smoothness (mean):                     0.053  0.163
compactness (mean):                    0.019  0.345
concavity (mean):                      0.0    0.427
concave points (mean):                 0.0    0.201
symmetry (mean):                       0.106  0.304
```

Analytics Vidhya
Learn everything about analytics

# Why is standardisation required?

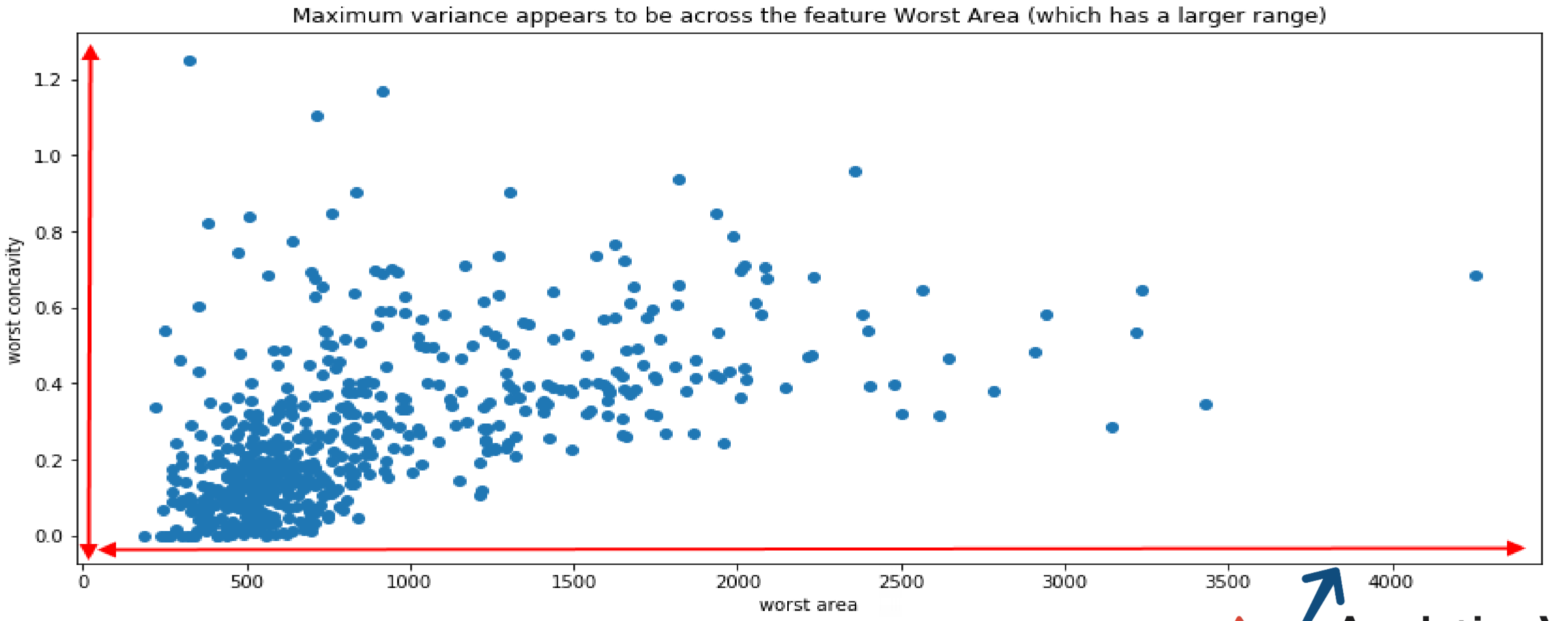
- A feature with a larger range will have a higher variance.



Analytics Vidhya
Learn everything about analytics

# Why is standardisation required?

- A feature with a larger range will have a higher variance.



Analytics Vidhya
Learn everything about analytics

# Why is standardisation required?

Standardisation is used so that each feature contributes equally to the PCA algorithm.

# How to standardise a value?

Mathematically, the standardised value of a value x is given by:

$$x_{new} = \frac{x - mean(x)}{std\ dev(x)}$$

# Properties of Standardised Data

Resultant features obtained after standardisation have the following properties:

- Distributed with mean = 0
- Distributed with variance = 1

The data is said to be **column-standardised**.

Thank
You!

Analytics Vidhya
Learn everything about analytics