



Dimensionality Reduction

Dimensionality Reduction ?

Reducing dimensions from higher to lower.

Dimensionality Reduction ?

Reducing dimensions from higher to lower.

But **why** dimension needs to be reduced ?

Dimensionality Reduction ?

Reducing dimensions from higher to lower.

But **why** dimension needs to be reduced ?

- To reduce complexity of model
- To prevent overfitting
- For better visualization

Dimensionality Reduction ?

Reducing dimensions from higher to lower.

But **why** dimension needs to be reduced ?

- To reduce complexity of model
- To prevent overfitting
- For better visualization

So, **how** to select important features ?

Dimensionality Reduction ?

Reducing dimensions from higher to lower.

But **why** dimension needs to be reduced ?

- To reduce complexity of model
- To prevent overfitting
- For better visualization

So, **how** to select important features ?

Filter Method

Calculates correlation coefficient score of each features and select based on score.

Wrapper Method

Predictive model is used to evaluate a combination of features and assign a score based on model accuracy

Embedded Method

Learn which features best contribute to the accuracy of the model

Dimensionality Reduction ?

Reducing dimensions from higher to lower.

But **why** dimension needs to be reduced ?

- To reduce complexity of model
- To prevent overfitting
- For better visualization

So, **how** to select important features ?

Filter Method

Calculates correlation coefficient score of each features and select based on score.

Wrapper Method

Predictive model is used to evaluate a combination of features and assign a score based on model accuracy

Embedded Method

Learn which features best contribute to the accuracy of the model

So, **what** are the techniques of Dimensionality reduction ?

Dimensionality Reduction Technique's

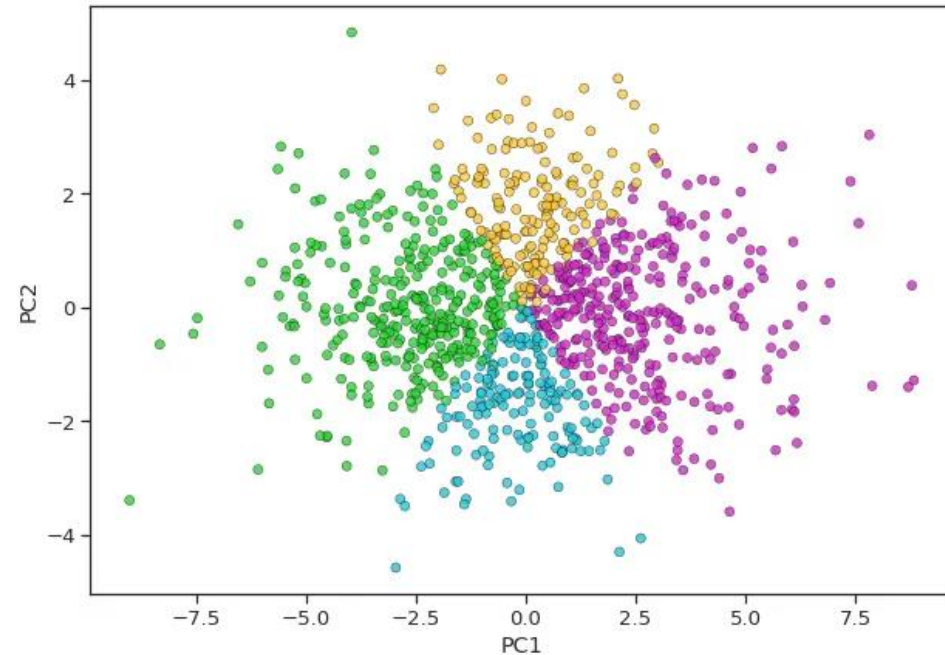
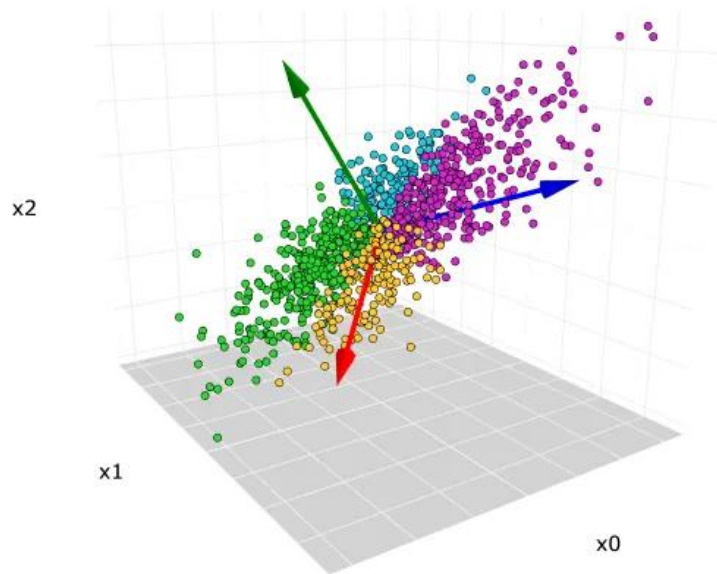
- PCA (Principal Component Analysis)
- LDA (Linear Discriminant Analysis)
- TSNE (T-distributed Stochastic Neighbor Embedding)
- UMAP (Uniform Manifold Approximation and Projection)
- SVD (Singular Value Decomposition)

Principal Component Analysis

Principal Component Analysis

PCA reduces higher dimensional data into lower dimensional space by identifying a set of new uncorrelated variable called principal components.

Principal Components are determined by computing eigenvectors of the covariance matrix.



Principal Component Analysis

PCA reduces higher dimensional data into lower dimensional space by identifying a set of new uncorrelated variable called principal components.

Principal Components are determined by computing eigenvectors of the covariance matrix.

Let's Perform PCA using “**US Arrests**” dataset on **R**

	Murder	Assault	UrbanPop	Rape
Alabama	13.2	236	58	21.2
Alaska	10.0	263	48	44.5
Arizona	8.1	294	80	31.0
Arkansas	8.8	190	50	19.5
California	9.0	276	91	40.6
Colorado	7.9	204	78	38.7

Principal Component Analysis

PCA reduces higher dimensional data into lower dimensional space by identifying a set of new uncorrelated variable called principal components.

Principal Components are determined by computing eigenvectors of the covariance matrix.

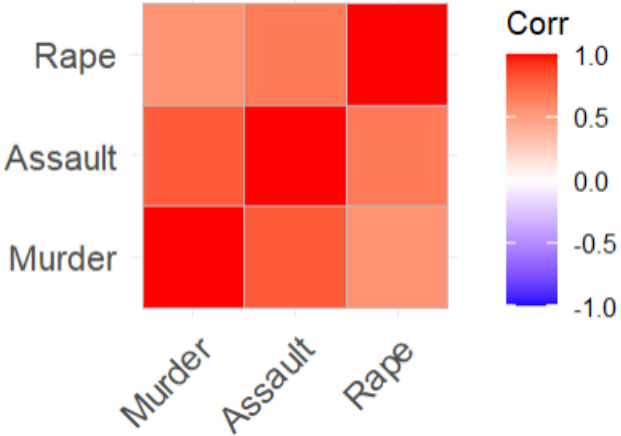
Let's Perform PCA using “**US Arrests**” dataset on **R**

	Murder	Assault	UrbanPop	Rape
Alabama	13.2	236	58	21.2
Alaska	10.0	263	48	44.5
Arizona	8.1	294	80	31.0
Arkansas	8.8	190	50	19.5
California	9.0	276	91	40.6
Colorado	7.9	204	78	38.7

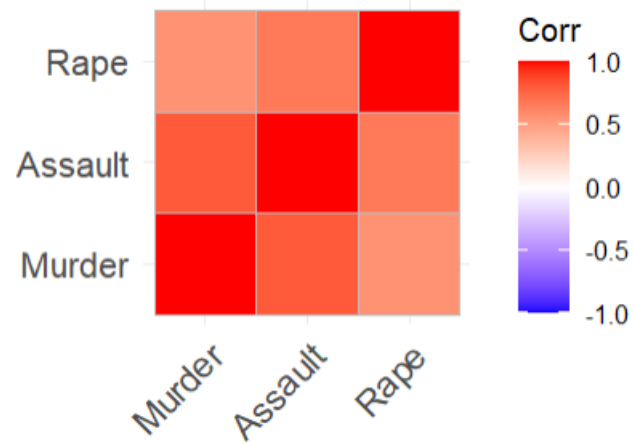
1. Data Normalization: Scale each of the variables to have a mean of 0 and a standard deviation of 1.

	Murder	Assault	Rape
Alabama	1.24256408	0.7828393	-0.003416473
Alaska	0.50786248	1.1068225	2.484202941
Arizona	0.07163341	1.4788032	1.042878388
Arkansas	0.23234938	0.2308680	-0.184916602
California	0.27826823	1.2628144	2.067820292
Colorado	0.02571456	0.3988593	1.864967207

2. Calculate the correlation matrix for the scaled variables.



2. Calculate the correlation matrix for the scaled variables.



3. Selection of Principal Components

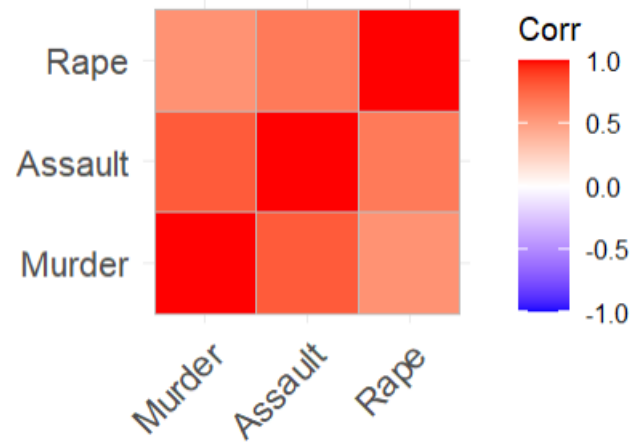
Importance of components:

	PC1	PC2	PC3
Standard deviation	1.5358	0.6768	0.42822
Proportion of Variance	0.7862	0.1527	0.06112
Cumulative Proportion	0.7862	0.9389	1.00000

We select PC1 to be retained as the Eigenvalue (SD) is ≥ 1

So, we use PC1 as the “**criminality score**” to represent 3 features using PCA method

2. Calculate the correlation matrix for the scaled variables.



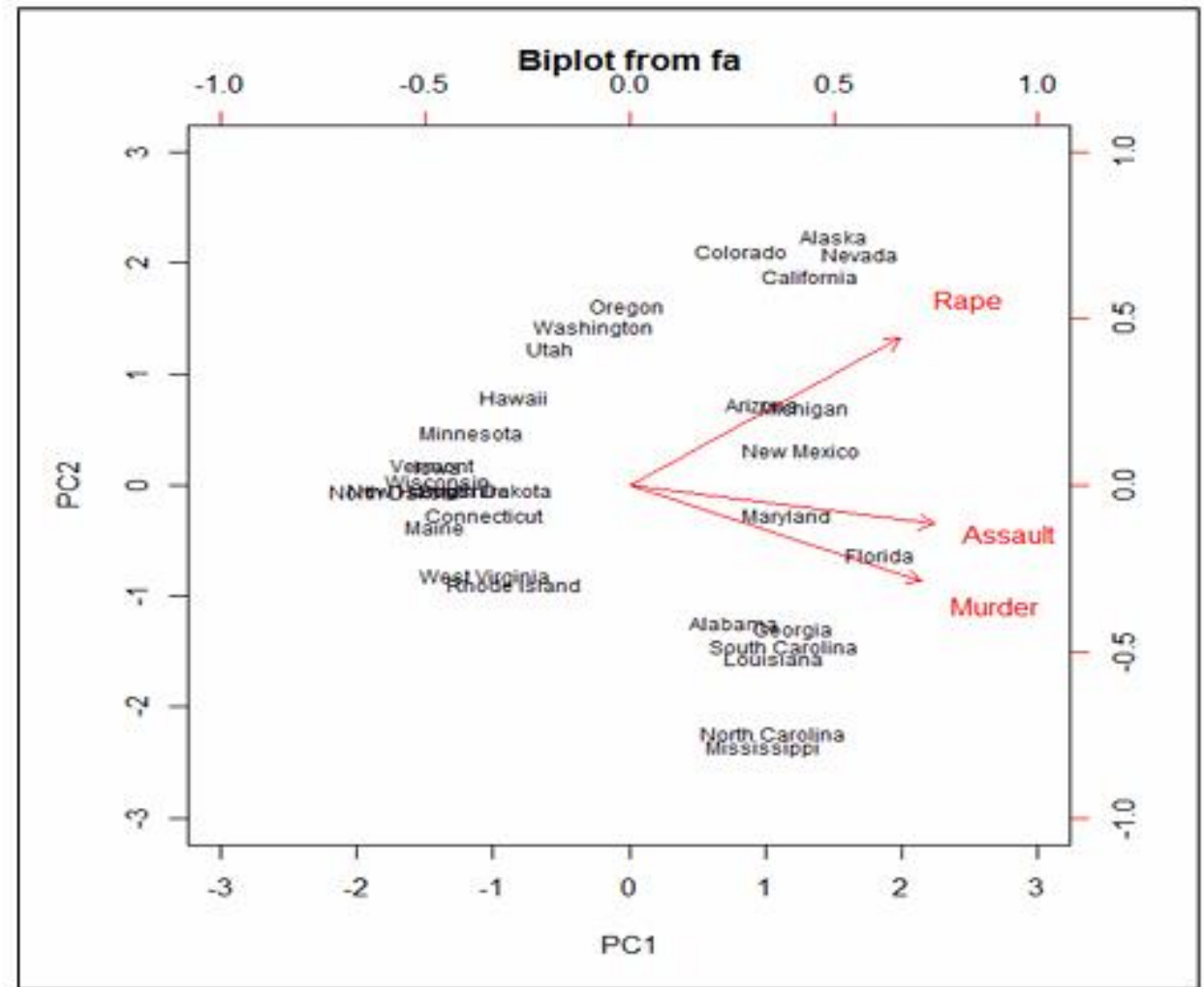
3. Selection of Principal Components

Importance of components:

	PC1	PC2	PC3
Standard deviation	1.5358	0.6768	0.42822
Proportion of Variance	0.7862	0.1527	0.06112
Cumulative Proportion	0.7862	0.9389	1.00000

We select PC1 to be retained as the Eigenvalue (SD) is ≥ 1

So, we use PC1 as the “**criminality score**” to represent 3 features using PCA method



THANK YOU