Dimensionality Reduction

Augmented reality technology has existed mostly as a Smartphone novelty --see yelp's "monocle" feature on their mobile app or the Layer augmented reality browser-- but the technology may soon become more integrated into our daily lives through the use of media-enhanced glasses. Most visibly, Google seems to be making the move toward augmented reality glasses. The New York Times recently reported that "The glasses will have a low-resolution built-in camera that will be able to monitor the world in real time and overlay information about locations, surrounding buildings and friends who might be nearby, according to the Google employees."

Dimensionality Reduction

- used to reduce number of feature
- Two types:

1. Feature Selection (keeps a subset of original features)
2. Feature Extraction (creates new one)

Advantages:

- less space is required
- less computation time
- takes care of multicollinearity by removing redundant features

Feature Selection

By only keeping the most relevant variables from the original dataset

- **Backward Feature Elimination** (Start removing one variable each time and check performance)
- **Forward Feature Selection** (Start with one variable and see which is best)

Feature Selection

Steps:

- starts with the evaluation of each individual feature and choose one that results in the best
- best depends on the chosen criteria e.g: scoring='accuracy'
- Next, all possible combinations of the first feature (from step1) and a second feature is selected based on evaluation
- this goes on until predefined number of feature is selected

Feature Extraction

1. **PCA (Principal Component Analysis)**
2. **LDA (Linear Discriminant Analysis)**

Principal Component Analysis (PCA)¶

- its a projection of a higher-dimensional space into a lower dimensional space while preserving as much information as possible
- Principal Components are linear combination of original set
- unsupervised algorithm as it ignores Dependent Variable and goal is to find the direction
- remove feature that explains less % of variance

PCA : Explained

- **First Level**:

1. Some wines bottle
2. Describe each wine by its colour, by how strong it is, by how old it is, and so on
3. Many of the characteristic will measure same property so redundant
4. Can summarize each wine with fewer characteristics

- **Second Level:** (So PCA checks what characteristics are redundant and discards them)

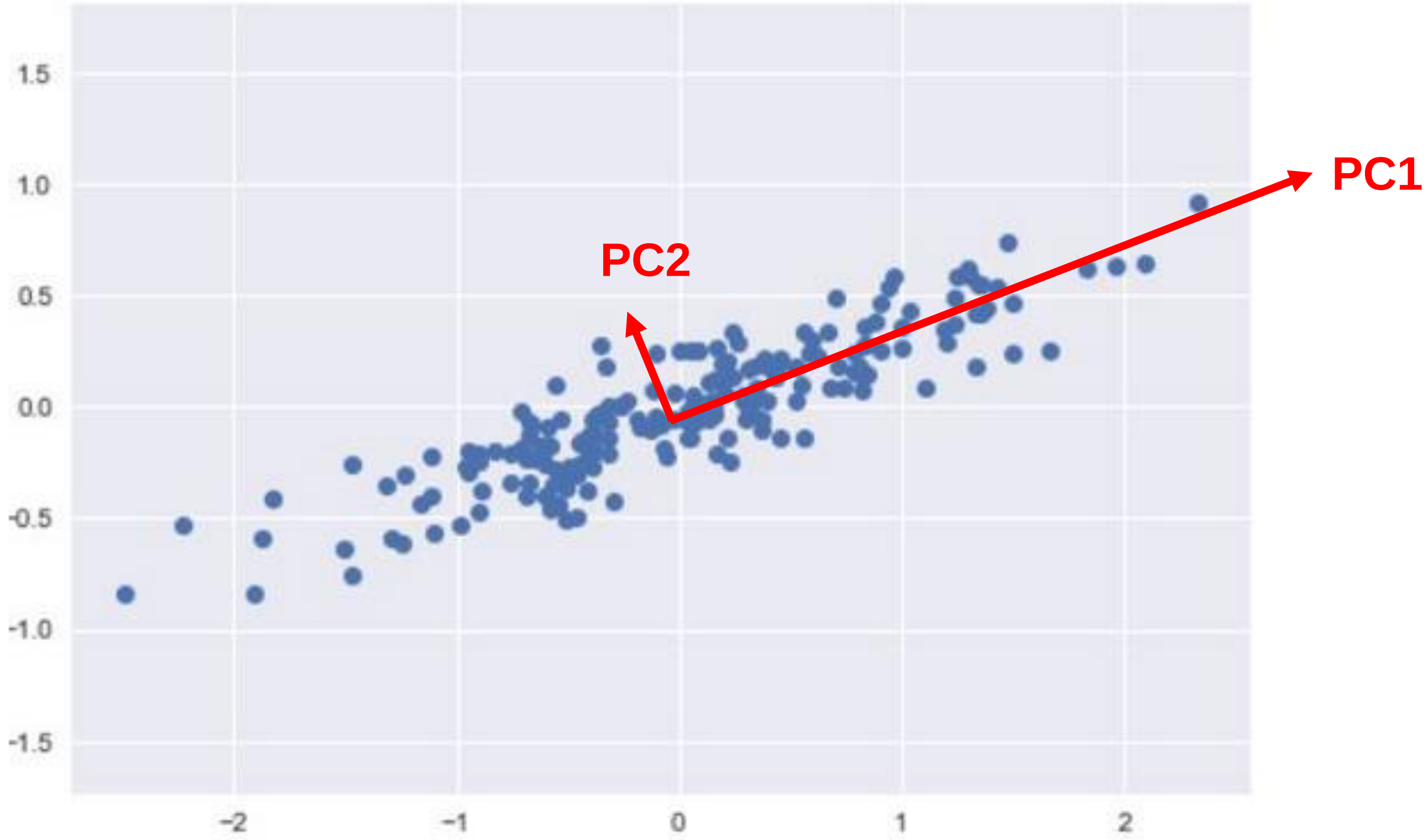
1. Nope
2. Constructs new characteristics using old ones

When should I use PCA?

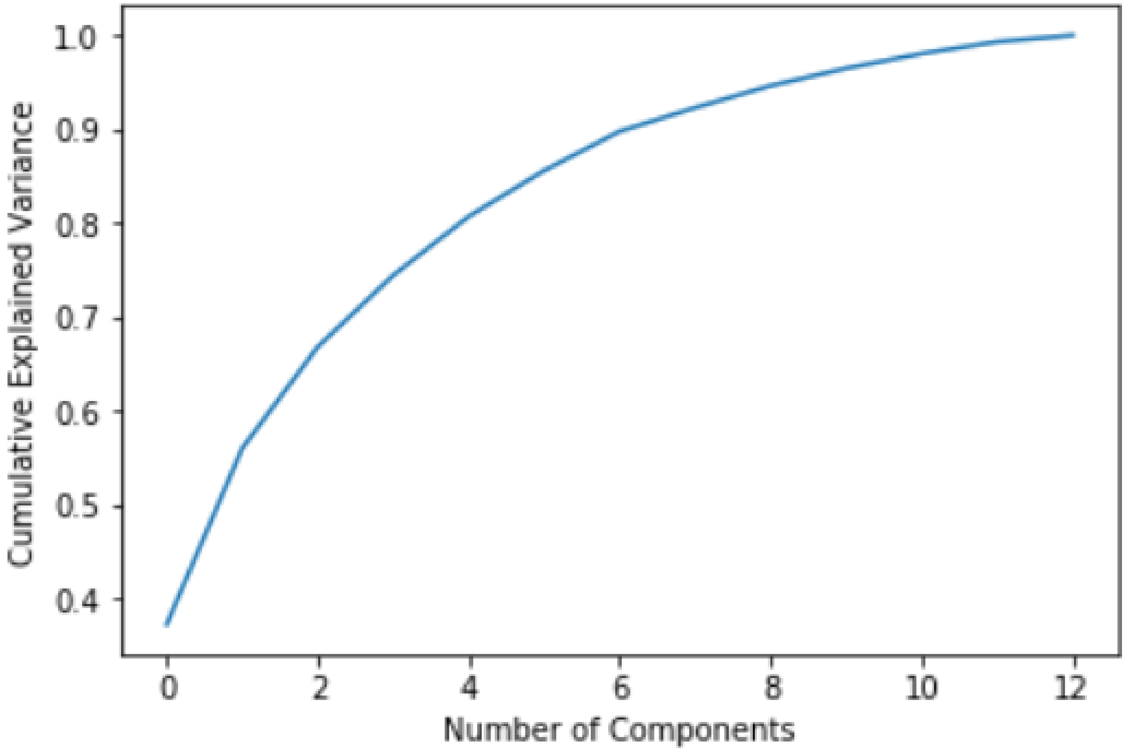
- Want to reduce the no. of variables, but don’t know which variable can be removed completely

- Want to ensure that variables are independent of one another?

- If making the independent variable less interpretable is fine?


1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
−2
−1
0
1
2
PC1
PC2

Choosing no. of Components



PCA

Issues:

- Measurements from all of the original variables are used in the projection to the lower dimensional space
- Only linear relationships are considered
- Do not consider the potential multivariate nature of the data structure (higher order interaction between variables)

PCA

How to address Issues:

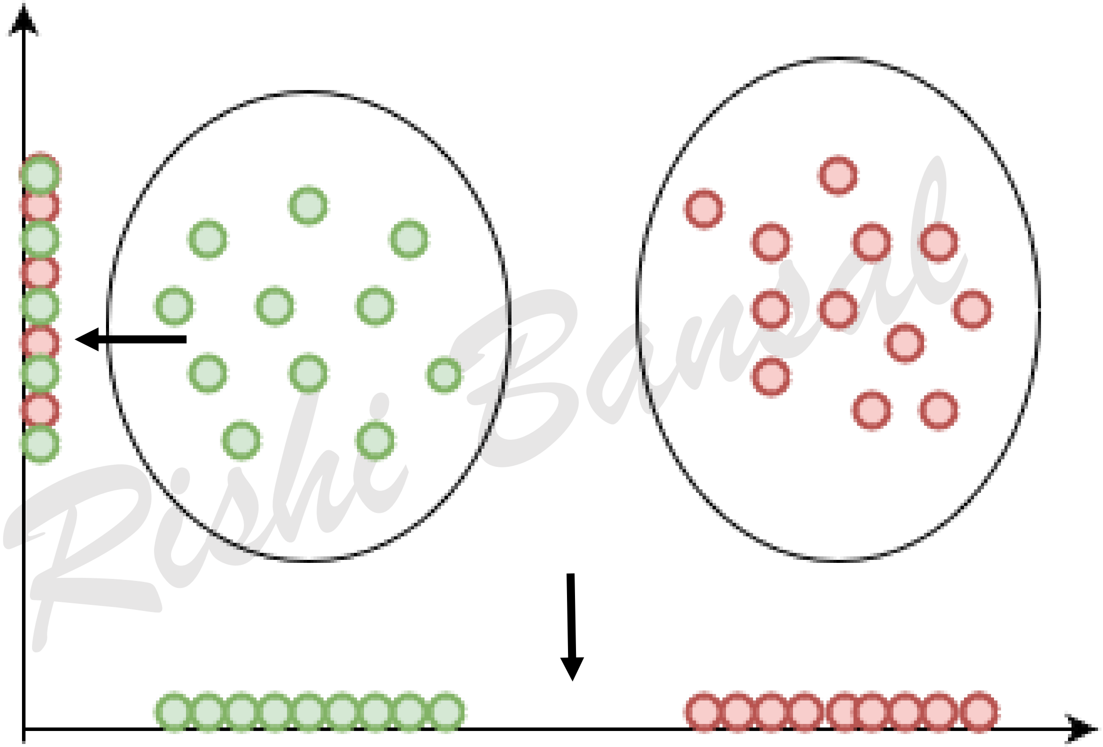
- Feature Selection Techiniques can be used
- Kernel PCA - to embed nonlinear relationships into a lower dimensional space
- Random Forest or Decision Tree. Random Forest derive Gini- or permutation-based measures of feature importance.

Advantage of PCA

- Improves performance of Algorithm(less feature)
- Help in removing correlated feature
- Reduces overfitting
- Reduce Training time

LDA

- Supervised as it considers Dependent Variable



- Maximizing the component axes for class-separation