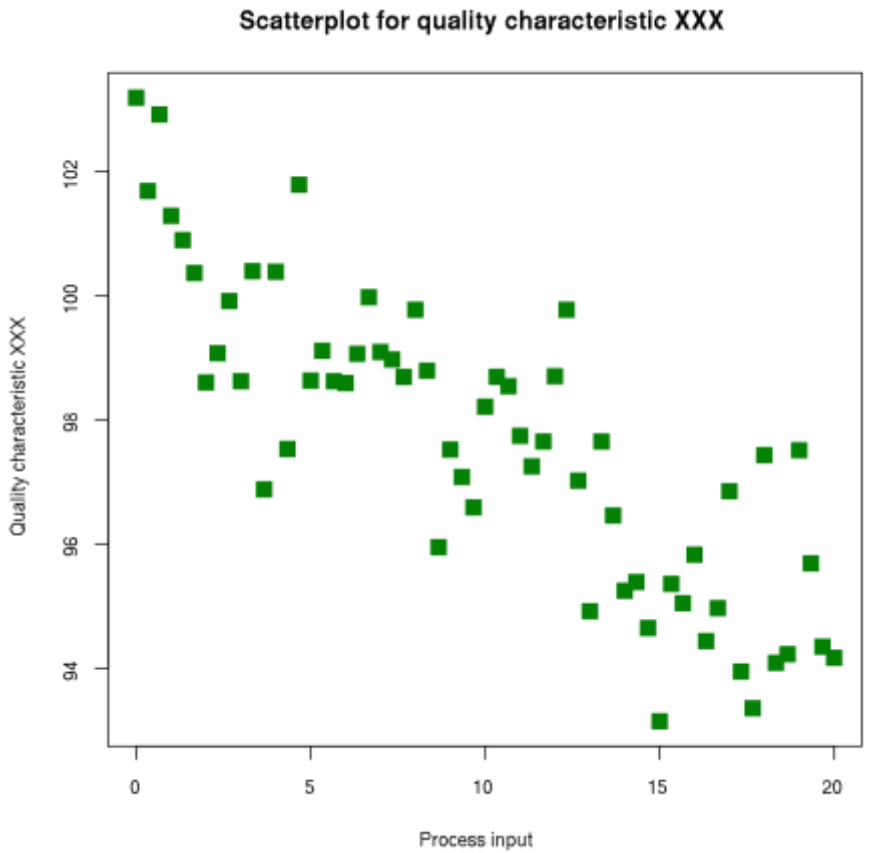
Data Visualization

Dr. Kalidas Y., IIT Tirupati

Scatter Plot

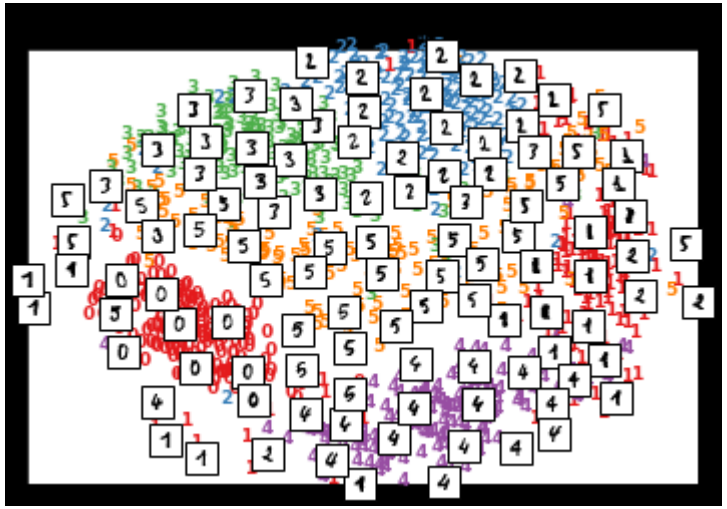
Plot – An attribute vs Target

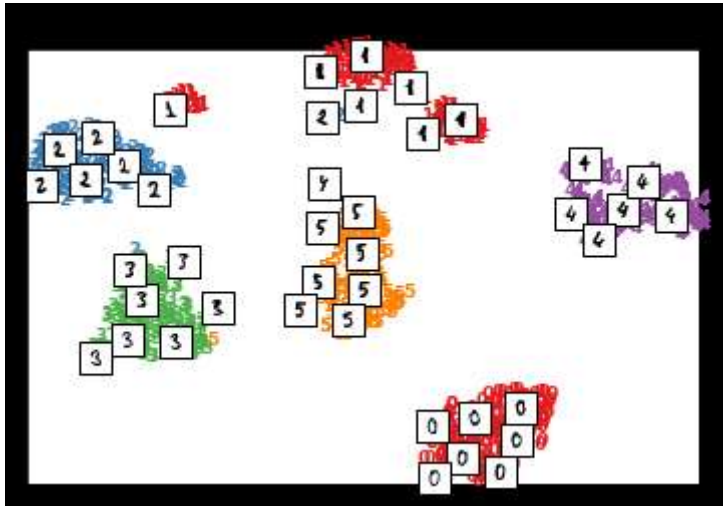


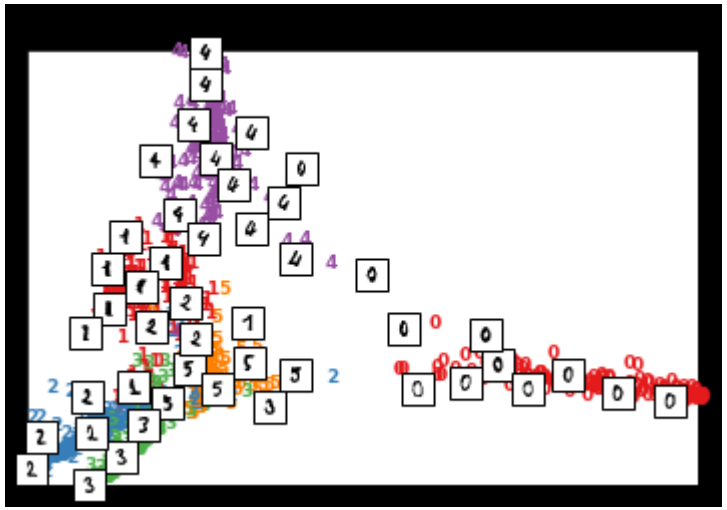
Visualize a high dimensional data set?

- Project it onto 2D plat and visualize!

- How to ‘project’?







PCA – Projection on to 2D

- PCA(n_components=2)
- PCA – fit(X)
- X_pca = PCA.transform(X)

- now X_pca is a set of 2D points, we can plot them and visualize

127) key phrase… “MDS” Multi Dimensional Scaling

- Formulation idea –
  ‘*pairwise distances be same before and after transformation*’
- Distance between points and dot product
  - $dist(X[i], X[j]) = \sqrt{\sum_{k=0}^{d-1}(X[i][k] - X[j][k])^2}$
  - $|a - b|^2 = |a|^2 + |b|^2 - 2\, a\, b \cos(\theta)$
  - $dist(X[i], X[j])^2 = |X[i]|^2 + |X[j]|^2 - 2\ (X[i] \cdot X[j])$
  - If all vectors are unit vectors, then
  - $d(X[i], X[j])^2 \propto 1 - k(X[i], X[j])$
  - Where k(a,b) is some similarity function between two vectors a and b
- We need to maintain *almost identical* distances between points before and after transformation
- Let
  - Input $X[i]$ be k dimensional
  - $W_{k\times 2}$ be the transformation matrix
  - Determine transformed coordinates, $Z = X \times x$
  - Dot products should remain same!
  - $X\, X^T \approx Z\, Z^T$
  - Loss function, $L(W) = |XX^T - XWW^TX|^2$ minimize this function over ‘W’ parameters

128) key phrase… “TSNE”
t-test based stochastic neighbourhood embedding