(1) Silhoutte

Calculates consistency of clusters. 🡪 Measures like own cluster (a) vs other clusters (b).

s = (b-a)/max(a,b), -1 <= s <= 1   
[**https://scikit-learn.org/stable/auto\_examples/cluster/plot\_kmeans\_silhouette\_analysis.html**](https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html)

Silhouette coefficients (as these values are referred to as) near +1 indicate that the sample is far away from the neighboring clusters. A value of 0 indicates that the sample is on or very close to the decision boundary between two neighboring clusters and negative values indicate that those samples might have been assigned to the wrong cluster.

Also from the thickness of the silhouette plot the cluster size can be visualized. The silhouette plot for cluster 0 when n\_clusters is equal to 2, is bigger in size owing to the grouping of the 3 sub clusters into one big cluster. However when the n\_clusters is equal to 4, all the plots are more or less of similar thickness and hence are of similar sizes as can be also verified from the labelled scatter plot on the right.

(2) TSNE

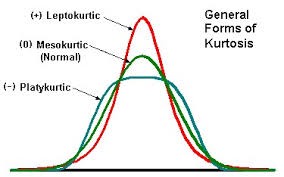
T = XW, X = values, W = weights,

(3) PCA

(4) components

**A feature**

(5) kurtosis



Analysis of the Tails. 🡪   
Mesokurtic = Tails similar to normal distr,   
Leptokurtic = long skinny tails, 🡪 chance of outliers   
Platykurtic = Flat. Top. Not much tail

<https://medium.com/dev-genius/skewness-and-kurtosis-in-data-science-aa795ba4b453>

A large kurtosis value often means that the tails of the distribution are getting more extreme values than the tail of the normal distribution. This may lead to a length of 6 or 7 standard deviations from the mean. Similarly, If the kurtosis value is very low, the tail of the distribution will be less lengthy than the tail of a normal distribution (less than 3 standard deviations).

A large value of kurtosis is often considered as riskier because data may tend to give an outlier value as an outcome with greater distance from the mean if applied to any machine learning algorithm.

(6) AIC/BIC Score

<https://www.methodology.psu.edu/resources/AIC-vs-BIC/>

AIC is an estimator of in-sample prediction error and thereby relative quality of statistical models for a given set of data

The formula for the Bayesian information criterion (BIC) is similar to the formula for AIC, but with a different penalty for the number of parameters. With AIC the penalty is 2k, whereas with BIC the penalty is ln(n) k.

Calculates

|  |  |
| --- | --- |
| BIC | 2k – 2\*ln(likelihood) |
| AIC | k\*ln(n) – 2\*ln(likelihood) |

K = # param, n = # samples, likelihood = p(x | theta, M), M = model, theta = param values

[**https://machinelearningmastery.com/probabilistic-model-selection-measures/**](https://machinelearningmastery.com/probabilistic-model-selection-measures/)

**SMALLER THE BETTER**

(7) variance/cum var

* **Variance** refers to the sensitivity of the learning algorithm to the specifics of the training data, e.g. the noise and specific observations. This is good as the model will be specialized to the data at the cost of learning random noise and varying each time it is trained on different data.

<https://machinelearningmastery.com/how-to-reduce-model-variance/>

Want to MAXIMIZE Variance for PCA. 🡪 Variance of the data in the low-dimensional representation is maximized 🡪 bubbles are far apart as possible.

(8) inertia

Tells you how far apart points within a cluster are. -🡪 **Small Inertia = Good!**

(9) that red line.

(10) Spherical vs Tied vs Diag vs Full

|  |  |  |  |
| --- | --- | --- | --- |
| diabetes | Normal | Kmeans | EM |
| Normal |  | **2 clusters** | **10 diag** |
| PCA | **16 components** | **2 Clusters!** | **(9) ~~10~~ full.** |
| ICA | 11 components  Has the lowest kurtosis | 2 Clusters | **10 full** |
| RP | **16 components** has lowest MSE | **3 clusters** has the lowest errors in depth and width | **10 full** |
| KPCA | **16 components**  Sigmoid or cosine  The errors both went under 0.025 which is better than poly. | 2 clusters | **10 full** |

|  |  |  |  |
| --- | --- | --- | --- |
| cancer | Normal | Kmeans | EM |
| Normal |  | **2 clusters.**  **3 cluster and 2 clusters have similar total errors. Seems like c1 and c2 is as confused as in c1 of clusterof2** | 10 full |
| PCA | 21 components  When down to 0 |  |  |
| ICA | 12 components  Least kurtosis before extreme.  Or 5/4/3/2 |  |  |
| RP | Only 30 components would reconstruct fully. |  |  |
| KPCA | Cosine kernel,  25 components  Went down to 0 |  |  |

|  |  |
| --- | --- |
| Normal |  |
| Pca |  |
| Ica |  |
| Rp |  |
| Kpca |  |
|  |  |

|  |  |
| --- | --- |
| Normal |  |
| Pca |  |
| Ica |  |
| Rp |  |
| Kpca |  |

EM

|  |  |
| --- | --- |
| Normal |  |
| Pca |  |
| Ica |  |
| Rp |  |
| Kpca |  |

|  |  |
| --- | --- |
| Normal |  |
| Pca |  |
| Ica |  |
| Rp |  |
| Kpca |  |