Schizophrenia Classification

Στατιστική Μοντελοποίηση και Αναγνώριση Προτύπων
Κατάρα Σωτηρία Μαρία
Μανάρα Χριστίνα

What is Schizophrenia?

- Mental Disorder
- Abnormal
- Strange Speech
- Decreased ability to understand reality

Other symptoms

- False Beliefs
- Hearing voices that do not exist
- Reduced social engagement

Causes

- Raised in a city
- Cannabis use during adolescence
- Age of parents
- Poor nutrition during pregnancy

What is Classification?

- Problem of identifying to which of a set of categories a new observation belongs to

Schizophrenia + Classification

- Diagnose subjects with schizophrenia based on multimodal features derived from their brain magnetic resonance imaging (MRI) scans
- High Dimensional Small Sample Size Data Problem

Purpose

- Training a model to classify a group of schizophrenic and healthy patients by using different algorithms for classification and ending up with the most efficient.

Training Data and Feature Selection

Training Data and Feature Selection

Training Data

FNC: Correlation values. They describe the correlation level between pairs of brain maps over time.

SBM: Standardized weights. They describe the expression level of ICA brain maps derived from gray-matter concentration .

Feature Selection

- Principal Component Analysis (PCA)
- From 410 dimensions to 80

Cross Validation

- Cross Validation is a very useful technique for assessing the effectiveness of your model, particularly in cases where you need to mitigate overfitting.
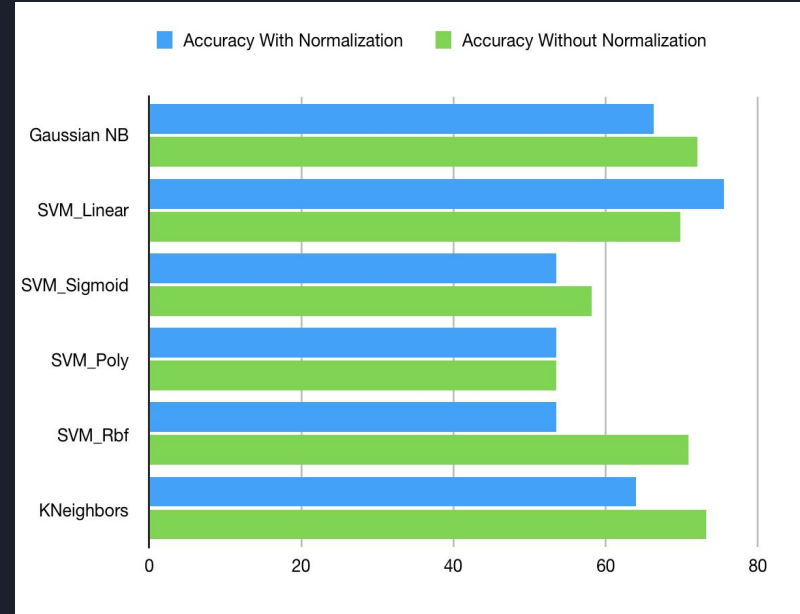
Classifiers

Classifiers

- K Nearest Neighbors (KNN)
- Gaussian Naive Bayes
- Support Vector Machine (SVM):
  - ✓ Linear kernel
  - ✓ Polynomial kernel
  - ✓ Radial basis function kernel
  - ✓ Sigmoid kernel

Results

- Worst accuracy up to 53.49% in SVM Sigmoid, Poly and Rbf kernel with normalization
- Worst accuracy up to 53.4% in SVM Poly kernel without normalization
- Best accuracy up to 75.58% in SVM Linear kernel with normalization
- Best accuracy up to 73.26% in k-NN without normalization

Conclusion

- Can not be achieved higher accuracy because of the problem of "High Dimensional Small Sample Size Data ".
- The number of samples does not suffice in order to achieve better training results for the classifiers.
- Don't normalize the data, unwittingly giving some features more importance than others so SVM doesn't work well.
- Non-normalization is better for K-NN, since the data needs to be more outdated for better performance.

Thank you!