

The goal of this particular project is to evaluate a set of classifiers(KNN, SVM, Decision Trees) based on their:

1. Sensitivity
2. Specificity
3. Accuracy

Dataset

The dataset that was used in this project was imported from Donald Bren School of Information & Computer Sciences and includes characteristics for breast-cancer diagnosis.

(<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>)

The dataset consists of 699 instances, each of those has one of 2 possible classes, either benign (denoted with the value of 2) or malignant (denoted with the value of 4). Each instance has also 9 attributes, scaled in [1,10] plus one more attribute which is the class's label (benign, malignant).

Missing attributes, which at the initialization step were denoted with the value "?" and were 16 in total, were replaced with the most frequent value.

Also, the distribution of these classes is:

Benign: 458 (65.5%)

Malignant: 241 (34.5%)

The results that were collected after using a variety of parameters for each classifier are the following:

#	Attribute	Domain
1	Sample code number	Id number
2	Clump Thickness	1 - 10
3	Uniformity of Cell Size	1 - 10
4	Uniformity of Cell Shape	1 - 10
5	Marginal Adhesion	1 - 10
6	Single Epithelial Cell Size	1 - 10
7	Bare Nuclei	1 - 10
8	Bland Chromatin	1 - 10
9	Normal Nucleoli	1 - 10
10	Mitoses	1 - 10
11	Class:	1 - 10

Classifiers

Classifiers that were implemented successfully are the following: **K-Nearest Neighbors, Support Vector Machines, Decision Trees**

Execution

In order to classify the data, type `main.m data.xlsx` following the name of the classifier you wish to evaluate. For example:

```
main.m data.xlsx svm to evaluate SVM
```

Results

Classifier	Accuracy	Sensitivity	Specificity
SVM	0.97	0.99	0.96
KNN	0.97	0.97	0.97
Decision Tree	0.95	0.97	0.94

According to the results above, the best overall classifier seems to be SVM. This conclusion arises by comparing the classifiers' Sensitivity, since it's important to maximize the True Positive percentage, which is the model's capability of correctly predicting the tumor's state.