

Faculty of physical-mechanical engineering, Systems Engineering and Informatics
Universidad Industrial de Santander “UIS”

Prostatex Dataset - experiments

v. 1.0.0

Yesid Gutiérrez
`yesid.gutierrez@correo.uis.edu.co`

November 14, 2018



Introduction

Introduction

Classification Experiments

Dataset information

Dataset integrity

First experiment

First experiment-Sample

First experiment-Results

Second experiment

Second experiment-Sample

Second experiment-Results

Third experiment

Third experiment-Sample

Third experiment-Results

Fourth experiment

Fourth experiment-Sample

Fourth experiment-Results

Content II

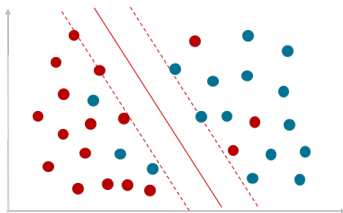


References

References



At this Assignment we are going to perform different experiments with our dataset. The experiments will be focused in two different techniques Classification experiments and Rigid Registration experiments.



Classification Experiments

Dataset information



The dataset contains k-images of 203 Patients, each k-image is composed by 16 slices, 128 rows and 128 columns for a shape of (16, 128, 128).

The dataset is divided into 4 different regions of the prostate (PZ peripheral zone , TZ transitional zone, SV seminal vesicle, AS anterior fibromuscular stroma).

Each patient has 1 or more findings that includes the ijk column (row,column,slice) of study and the class for the finding (The patient has/ has not cancer). The total number of findings is 335.

Classification Experiments

Dataset integrity



The dataset has 3 problems at the data, the slices at the k-images don't match with the (row,column,slice) at the findings. Since the total data of the Dataset is 335 and there were 3 problems, The dataset have $3/335 = .0.89552$ % of the data corrupted. This data is going to be skipped in the experiments.

```
checkout_Kings_problems()
```

There were 3 problems

It was the 0.89552% of the dataset

	ProxID	ijk	State	zone
34	ProstateX-0025	89 67 24	True	PZ
159	ProstateX-0105	63 78 17	True	SV
248	ProstateX-0154	57 72 16	True	SV

Classification Experiments

First experiment



At the first experiment I'm going to perform the following actions:

- ▶ I am going to select all the rows, all the columns and the slice specified at (ijk column) in the findings for each k-images of the patients .
- ▶ I'm going to separate the dataset into the four different zones of the prostate in order to study them separately.
- ▶ Classifiers, For this experiment I am going to use the following classifiers: (GaussianNB, SVM, Decision Tree, Random Forest with 20 trees, KNeighbors with 20 neighbors)

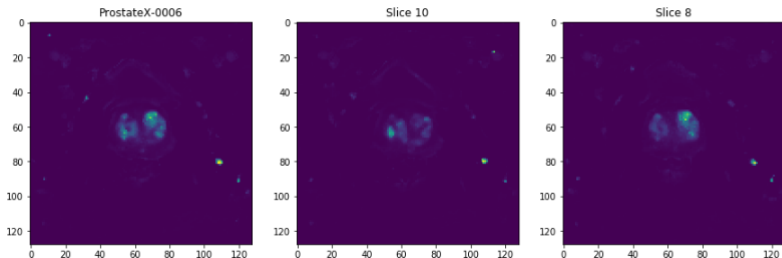
Classification Experiments

First experiment-Sample



The following figure is going to show an example of the data at training.

```
el estado es: False  
El voxel corresponde a (ijk): 74 70 9  
Las dimensiones del voxel son: (16, 128, 128)
```



Classification Experiments

First experiment-Results



the results of the first experiment are shown in the next figure:

```
display(classifiers_table)
```

	Classifier name	Prostate zone	Samples	Score	TPR	P	TNR	N
0	GaussianNB	PZ	75/97	0.773196	0.222222	18	0.898734	79
1	SVM-SVC	PZ	79/97	0.814433	0	18	1	79
2	Decision Tree	PZ	77/97	0.793814	0.222222	18	0.924051	79
3	Random Forest-20	PZ	79/97	0.814433	0.055556	18	0.987342	79
4	KNeighbors-20	PZ	79/97	0.814433	0	18	1	79
5	GaussianNB	AS	12/28	0.428571	0.625	16	0.166667	12
6	SVM-SVC	AS	15/28	0.535714	0.875	16	0.083333	12
7	Decision Tree	AS	13/28	0.464286	0.625	16	0.25	12
8	Random Forest-20	AS	12/28	0.428571	0.375	16	0.5	12
9	KNeighbors-20	AS	16/28	0.571429	1	16	0	12
10	GaussianNB	TZ	37/42	0.880952	0	5	1	37
11	SVM-SVC	TZ	37/42	0.880952	0	5	1	37
12	Decision Tree	TZ	27/42	0.642857	0.4	5	0.675676	37
13	Random Forest-20	TZ	35/42	0.833333	0.2	5	0.918919	37
14	KNeighbors-20	TZ	37/42	0.880952	0	5	1	37

Where TPR is the true positive rate, TNR is the true negative rate, P is the total of positive classes at testing and N is the total of negative classes at testing.

Classification Experiments

Second experiment



At the second experiment I'm going to perform the following actions:

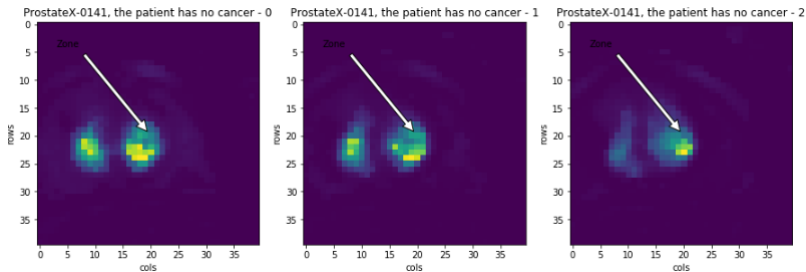
- ▶ I'm going to create sub voxels of 3 slices with 40 rows and 40 columns (3,40,40). the subvoxel is going to be centrated at the specified ijk column at the findings
- ▶ I'm going to separate the dataset into the four different zones of the prostate in order to study them separately.
- ▶ Classifiers, For this experiment I am going to use the following classifiers: (GaussianNB, SVM, Decision Tree, Random Forest with 20 trees, KNeighbors with 20 neighbors)

Classification Experiments

Second experiment-Sample



The following figure is going to show an example of the data at training.



Classification Experiments

Second experiment-Results



the results of the second experiment are shown in the next figure:

classifiers_table

	Classifier name	Prostate zone	Samples	Score	TPR	P	TNR	N
0	GaussianNB	PZ	52/97	0.536082	0.388889	18	0.56962	79
1	SVM-SVC	PZ	79/97	0.814433	0	18	1	79
2	Decision Tree	PZ	69/97	0.71134	0.222222	18	0.822785	79
3	Random Forest-20	PZ	74/97	0.762887	0.111111	18	0.911392	79
4	KNeighbors-20	PZ	79/97	0.814433	0	18	1	79
5	GaussianNB	AS	14/28	0.5	0.5	16	0.5	12
6	SVM-SVC	AS	17/28	0.607143	0.5625	16	0.666667	12
7	Decision Tree	AS	10/28	0.357143	0.5625	16	0.0833333	12
8	Random Forest-20	AS	19/28	0.678571	0.5625	16	0.833333	12
9	KNeighbors-20	AS	15/28	0.535714	0.9375	16	0	12
10	GaussianNB	TZ	37/42	0.880952	0	5	1	37
11	SVM-SVC	TZ	37/42	0.880952	0	5	1	37
12	Decision Tree	TZ	36/42	0.857143	0.2	5	0.945946	37
13	Random Forest-20	TZ	38/42	0.904762	0.2	5	1	37
14	KNeighbors-20	TZ	37/42	0.880952	0	5	1	37

Where TPR is the true positive rate, TNR is the true negative rate, P is the total of positive classes at testing and N is the total of negative classes at testing.

Classification Experiments

Third experiment



At the third experiment I'm going to perform the following actions:

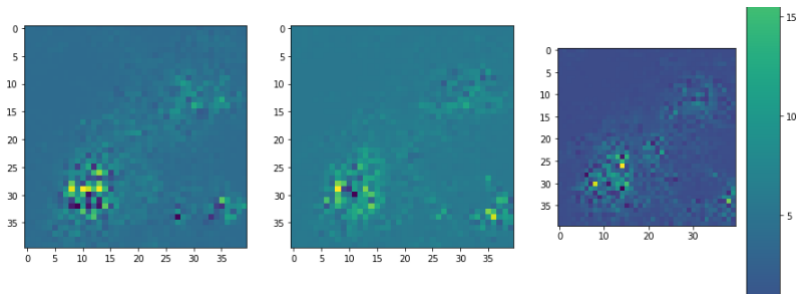
- ▶ I'm going to create sub voxels of 3 slices with 40 rows and 40 columns (3,40,40). the subvoxel is going to be centrated at the specified ijk column at the findings.
- ▶ I'm going to create virtual data in order to improve the balance of the classes at training.
- ▶ I'm going to separate the dataset into the four different zones of the prostate in order to study them separately.
- ▶ Classifiers, For this experiment I am going to use the following classifiers: (GaussianNB, SVM, Decision Tree, Random Forest with 20 trees, KNeighbors with 20 neighbors)

Classification Experiments

Third experiment-Sample



The following figure is going to show an example of the virtual data created at training.



Classification Experiments

Third experiment-Results



the results of the third experiment are shown in the next figure:

classifiers_table

	Classifier name	Prostate zone	Samples	Score	TPR	P	TNR	N
0	GaussianNB	PZ	51/97	0.525773	0.333333	18	0.56962	79
1	SVM-SVC	PZ	64/97	0.659794	0.277778	18	0.746835	79
2	Decision Tree	PZ	68/97	0.701031	0.222222	18	0.810127	79
3	Random Forest-20	PZ	80/97	0.824742	0.166667	18	0.974684	79
4	KNeighbors-20	PZ	79/97	0.814433	0	18	1	79
5	GaussianNB	AS	14/28	0.5	0.5625	16	0.416667	12
6	SVM-SVC	AS	16/28	0.571429	0.375	16	0.833333	12
7	Decision Tree	AS	14/28	0.5	0.375	16	0.666667	12
8	Random Forest-20	AS	16/28	0.571429	0.5	16	0.666667	12
9	KNeighbors-20	AS	14/28	0.5	0.25	16	0.833333	12
10	GaussianNB	TZ	37/42	0.880952	0	5	1	37
11	SVM-SVC	TZ	25/42	0.595238	0.2	5	0.648649	37
12	Decision Tree	TZ	35/42	0.833333	0.2	5	0.918919	37
13	Random Forest-20	TZ	36/42	0.857143	0.2	5	0.945946	37
14	KNeighbors-20	TZ	38/42	0.904762	0.2	5	1	37

Where TPR is the true positive rate, TNR is the true negative rate, P is the total of positive classes at testing and N is the total of negative classes at testing.

Classification Experiments

Fourth experiment



At the fourth experiment I'm going to perform the following actions:

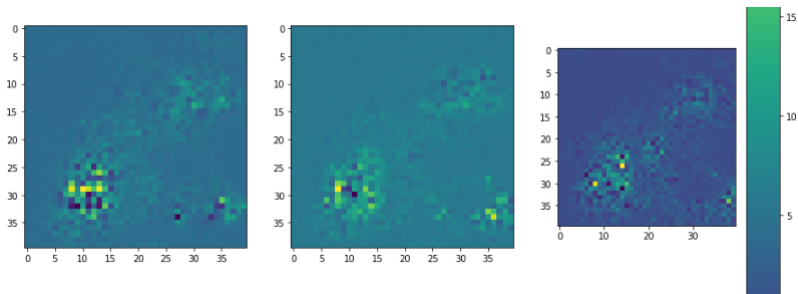
- ▶ I'm going to create sub voxels of 3 slices with 40 rows and 40 columns (3,40,40). the subvoxel is going to be centrated at the specified ijk column at the findings.
- ▶ I'm going to create virtual data in order to improve the balance of the classes at training.
- ▶ I'm going to create virtual data in order to improve the balance of the classes at testing.
- ▶ I'm going to separate the dataset into the four different zones of the prostate in order to study them separately.
- ▶ Classifiers, For this experiment I am going to use the following classifiers: (GaussianNB, SVM, Decision Tree, Random Forest with 20 trees, KNeighbors with 20 neighbors)

Classification Experiments

Fourth experiment-Sample



The following figure is going to show an example of the virtual data created at training and testing.



Classification Experiments

Fourth experiment-Results



the results of the fourth experiment are shown in the next figure:

classifiers_table

	Classifier name	Prostate zone	Samples	Score	TPR	P	TNR	N
0	GaussianNB	PZ	51/158	0.322785	0.0759494	79	0.56962	79
1	SVM-SVC	PZ	70/158	0.443038	0.151899	79	0.734177	79
2	Decision Tree	PZ	87/158	0.550633	0.35443	79	0.746835	79
3	Random Forest-20	PZ	106/158	0.670886	0.392405	79	0.949367	79
4	KNeighbors-20	PZ	79/158	0.5	0	79	1	79
5	GaussianNB	AS	14/32	0.4375	0.5	16	0.375	16
6	SVM-SVC	AS	20/32	0.625	0.375	16	0.875	16
7	Decision Tree	AS	18/32	0.5625	0.625	16	0.5	16
8	Random Forest-20	AS	17/32	0.53125	0.4375	16	0.625	16
9	KNeighbors-20	AS	18/32	0.5625	0.25	16	0.875	16
10	GaussianNB	TZ	37/74	0.5	0	37	1	37
11	SVM-SVC	TZ	57/74	0.77027	0.891892	37	0.648649	37
12	Decision Tree	TZ	38/74	0.527027	0.162162	37	0.891892	37
13	Random Forest-20	TZ	37/74	0.5	0	37	1	37
14	KNeighbors-20	TZ	38/74	0.513514	0.027027	37	1	37

Where TPR is the true positive rate, TNR is the true negative rate, P is the total of positive classes at testing and N is the total of negative classes at testing.



- ▶ https://simpleitk.github.io/SPIE2018_COURSE/basic_registration.pdf
- ▶ https://simpleitk.github.io/SPIE2018_COURSE/images_and_resampling.pdf
- ▶ <https://scikit-learn.org/stable/modules/svm.html>
- ▶ https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html
- ▶ <https://scikit-learn.org/stable/modules/svm.html>
- ▶ <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- ▶ <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>



Thanks for your attention!