

## Feature Vector Classification (Machine Learning)

Object identification by feature classification is an important final stage in many computer vision applications. This tutorial shows how to train and analyze the performance of a number of different classifications for the two class problem.

The commands `vrclasstt` and `vrclasscv` use the classification programs in the R statistical package

Yiting Xie and Anthony Reeves  
October, 2016

### Overview

#### **vrclasstt: classification with training set and test set**

1. Training a classifier:  
**vrclasstt** tr=train.csv of=results.csv cl=<classifier-type> om=model
2. Testing with a trained classifier:  
**vrclasstt** te=test.csv im=model cl=<classifier-type> of=results.csv
3. Training and testing a classifier  
**vrclasstt** tr=train.csv te=test.csv cl=<classifier-type> of=results.csv

#### **vrclasscv: classification analysis using cross validation**

**vrclasscv** tr=train.csv cl=<classifier-type> of=results.csv, roc=roc.png

Training, testing and response files are text files in csv format with a header



## Data file formats used by vrclasstt and vrclasscv

Main data files are in comma separated values (.csv) format with a header; each data row contains a labeled vector for classification

Feature Vector Format: (**train.csv** or **test.csv**)

<id<sub>1</sub>>,<id<sub>2</sub>>...<id<sub>n</sub>>, class, <fv<sub>1</sub>>,<fv<sub>2</sub>>, . . . <fv<sub>n</sub>>

Response Format: (**results.csv**)

<id<sub>1</sub>>,<id<sub>2</sub>>...<id<sub>n</sub>>, class, resp

Where

id: Identifier element

fv: Feature vector element

class: is the correct (truth) class

resp: is the numeric response from the classifier

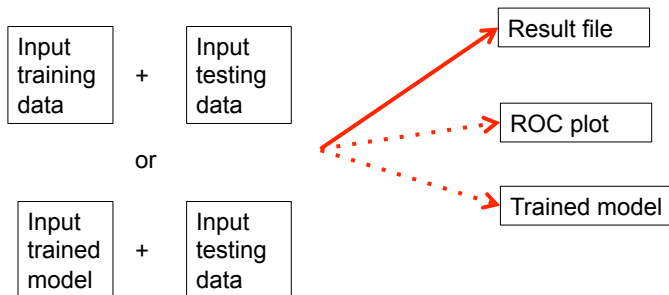
**model**: Trained classifier model in R internal format; contents depend upon classifier type



**roc.png**: ROC graphic in .png format.

Cornell University Vision and Image Analysis Group

## vrclasstt - Algorithm Details



Options:  
cl="knn" or "svmr" or  
"svmp" or "log" or "rf"

knn=K Nearest Neighbor (k=3);  
svmr=Support Vector Machine (RBF);  
svmp=Support Vector Machine (Polynomial);  
log=Logistic Regression;  
rf=Random Forest;



Cornell University Vision and Image Analysis Group

## vrclasstt – Training/Testing Example

vrclasstt tr=train.csv te=test.csv cl=log of=results.csv roc=roc.png

**train.csv:**

identifier	class	dmd	...
ID1	1	859.4	...
ID2	1	953.2	...
ID3	-1	995.5	...
...	...	...	...

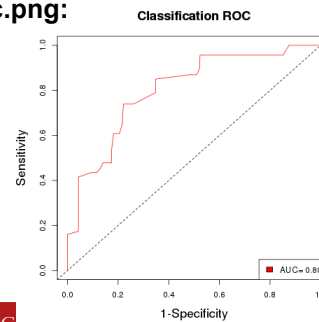
**test.csv:**

identifier	class	dmd	...
ID21	1	813.8	...
ID22	-1	996.0	...
ID23	-1	987.3	...
...	...	...	...

**results.csv:**

identifier	class	resp
ID21	1	0.72
ID22	-1	0.03
ID23	-1	0.25
...	...	...

**roc.png:**



Cornell University Vision and Image Analysis G

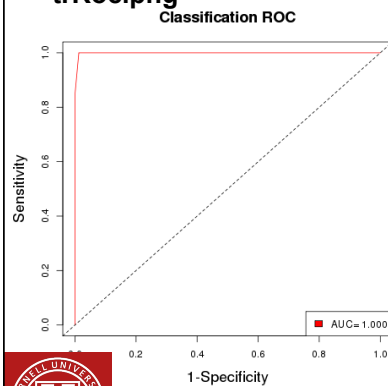
## vrclasstt – Testing Using Trained Model

**Command line:**

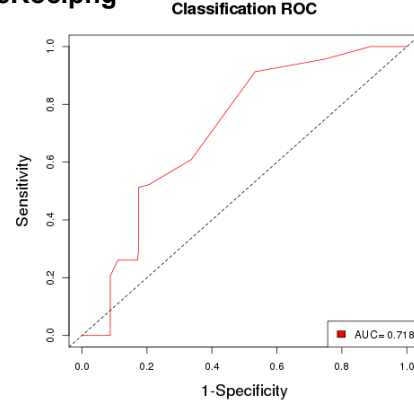
vrclasstt tr=train.csv om=model cl=svm of=trResults.csv roc=trRoc.png

vrclasstt im=model te=test.csv cl=svm of=teResults.csv roc=teRoc.png -e

**trRoc.png**



**teRoc.png**



Cornell University Vision and Image Analysis Group

## vrclasstt – Testing With Unknown Class Values

### Command line:

```
vrclasstt tr=train.csv om=model cl=svmp of=trainResults.csv
roc=trainRoc.png
```

```
vrclasstt im=model te=test.csv cl=svmp of=testResults.csv
```

### test.csv

identifier	class	dmd	...
ID21		813.8	...
ID22		996.0	...
ID23		987.3	...
...	...	...	...

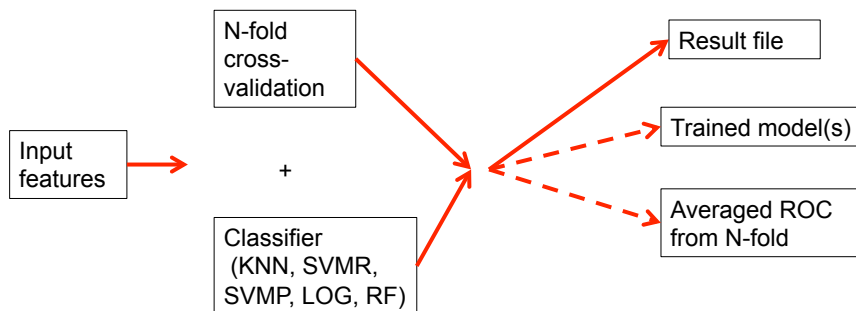
### testResults.csv

identifier	class	resp
ID21	NA	0.72
ID22	NA	0.03
ID23	NA	0.25
...	...	...



Cornell University Vision and Image Analysis Group

## vrclasscv - Algorithm Details



Options:  
n= N-fold N value;

Classifiers:  
knn=K Nearest Neighbor (k=3);  
svmr=Support Vector Machine (RBF);  
svmp=Support Vector Machine (Polynomial);  
log=Logistic Regression;  
rf=Random Forest;



Cornell University Vision and Image Analysis Group

## vrclasscv – Cross Validation Example

### Command line:

```
vrclasscv tr=train.csv cl="svm" of=results.csv roc=roc.png
```

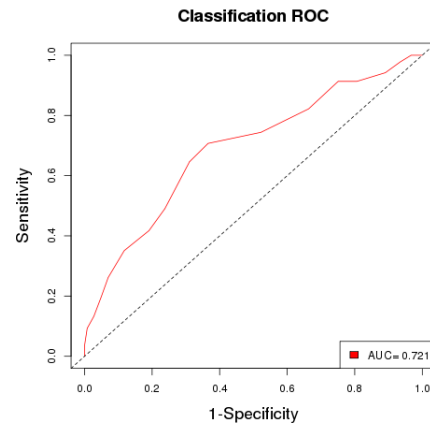
### train.csv:

identifier	class	gvol	...
ID1	1	186.8	...
ID2	1	253.5	...
ID3	-1	272.2	...
...	...	...	...

### results.csv:

identifier	class	resp
ID1	1	1.01
ID2	1	0.76
ID3	-1	0.35
...	...	...

### roc.png



Cornell University  
Vision and Image Analysis Group

VIA