Matters of Discussion

Performance Evaluation:

Evaluating classification performance

[Review]

Classification Performance - Evaluating Predictive Performance

Compiled By: Dr. Nilamadhab Mishra [(PhD- CSIE) Taiwan]

Performance measures

❖The performance of the developed model can be evaluated using **Confusion Matrix**

=== Confusion Matrix ===

a b <-- classified as

150 28 | a = tested_negative

32 51 | b = tested_positive

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

		Predicted Class	
		Class = **Negative**	Class =**Positive**
Actual Class	**Class = Negative [a]**	True Negative (TN)	False Positive (FP)
	Class = Positive [b]	False Negative (FN)	True Positive (TP)

Data set for building Confusion Matrix Example

A	P
1	1
1	1
1	0
1	0
0	1
0	1
0	0
0	0
1	1
1	1

		Predicted	
		1	0
Actual	1	4	2
	0	2	2

TP	FP
FN	TN

Performance measures

- Performance metrics

$$Sensitivity\ (\%) = \frac{TP}{TP+FN}\ X\ 100$$

$$Specificity\ (\%) = \frac{TN}{TN+FP}\ X\ 100$$

$$Precision\ (\%) = \frac{TP}{TP+FP}\ X100$$

$$Recall\ (\%) = Sensistivity\ (\%) = \frac{TP}{TP+FN}\ X\ 100$$

$$Accuracy\ (\%) = \frac{TP+TN}{TP+FP+TN+FN}\ X\ 100$$

Classifications - Classification methods

Decision Tree,

Naïve Bayes,

K-Nearest Neighbors

Already Discussed - ok

how to estimate the performance of those algorithms based on the measures.

Now we investigate the performance measure?

Performance measure for Naïve Bayes classification[class wise]

Detailed Accuracy by Class for Naïve Bayes classification algorithm

Detailed Accuracy by Class					
	TP Rate	Precision	Recall	F-Measure	Class
	97.3%	99.5%	97.3%	98.4%	FMD
	99.6%	97.7%	99.6%	98.6%	Blackleg
	99.1%	94.7%	99.1%	96.9%	Enteritis
	100%	97.9%	100%	98.9%	LSD
	96.9%	99.8%	96.9%	98.3%	Pneumonia
	98.0%	99.1%	98.0%	98.6%	Internal Parasite
	100%	99.6%	100%	99.8%	Anthrax
	96.9%	100%	96.9%	98.4%	Mastitis
Weighted Ave	98.5%	98.5%	98.5%	98.5%	

Comparison Result

Algorithms used	Correctly classified instances		Incorrectly classified instances		Time taken (second)
10 Fold Cross Validation Test Option					
	No.	Accuracy	No.	Accuracy	
Naive Bayes	3572	98.48 %	55	1.52%	0.02
J48	**3578**	**98.65%**	**49**	**1.35%**	**0.08**
JRip	3579	98.68%	48	1.32%	0.27
Percentage Split Test Option					
Naive Bayes	1070	98.35 %	18	1.65 %	0.02
J48	1069	98.25 %	19	1.75 %	0.06
JRip	1069	98.25 %	19	1.75%	0.03

Four common Test options

For both, **training** and **testing**, you need data. Those four options are commonly used.

1. Use training set:

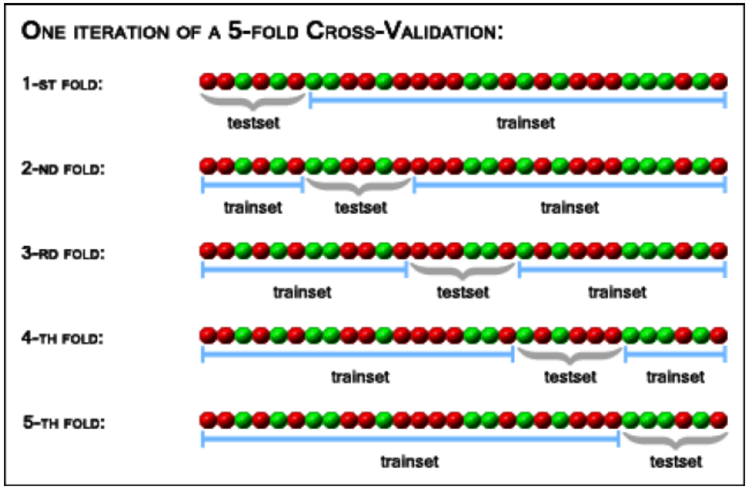
- Means you will test your knowledge on the same data you learned.
- Not very accepted because you can just make build your code to memorize the training instances (which will be in the test).
- Less degree of use for research.

2. Supplied test set:

- It is an external file that you can use as training set.
- It can be used when you want/need to test the algorithm's knowledge against a specific test set.

3. K-fold cross validation

- The training set is randomly divided into K disjoint sets of equal size where each part has roughly the same class distribution.
- You fold the data in 10 folds (for example) and repeat 10 (because it is 10-folds) the following process: Use 9 folds for training and leave 1 fold out for testing.



4. Percentage split:

- Splits the data and separates x% of the data for learning and the rest of it for testing.
- It is useful when your algorithm is slow.
- The best method to evaluate your classifier is to train algorithm with 67% of your training data and 33% to test your classifier.

Model performance for classification models

A classification model is a machine learning model which predicts a Y variable which is categorical:

1. Will the employ leave the organization or stay?
2. Does the patient have cancer or not?
3. Does this customer fall into high risk, medium risk or low risk?
4. Will the customer pay or default a loan?

A classification model in which the Y variable can take only 2 values is called a binary classifier.

CASE: Confusion matrix for customer class prediction

=== Confusion Matrix ===

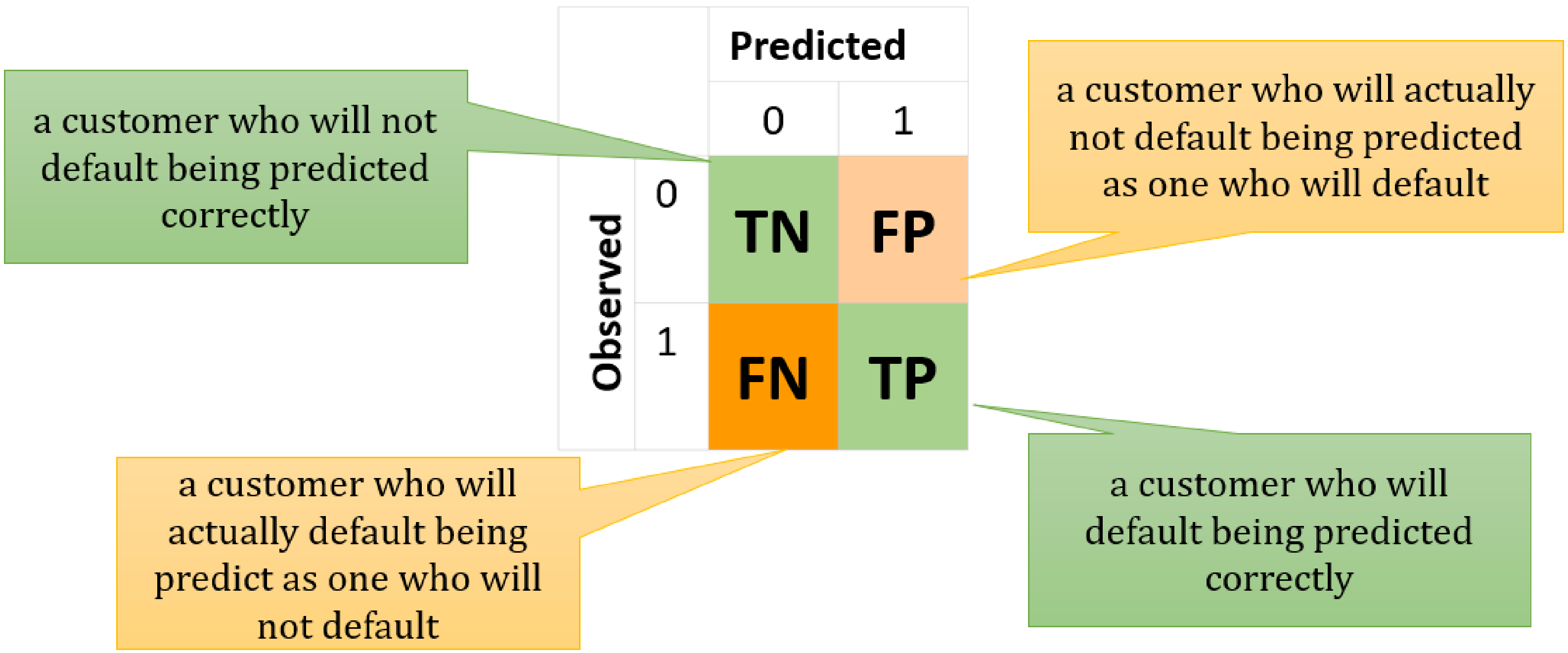
a b <-- classified as

150 28 | a = tested_negative

32 51 | b = tested_positive

TN= 150 ; FP = 28

FN= 32 ; TP = 51



model performance measure

1. Accuracy: = [TP+TN] / [TP+FP+TN+FN]

Accuracy is the number of correct predictions made by the model by the total number of records. The best accuracy is 100% indicating that all the predictions are correct.

TN= 150 ; FP = 28

2. Sensitivity or recall

FN= 32 ; TP = 51

Sensitivity (**Recall or True positive rate**) is calculated as the number of correct positive predictions divided by the total number of positives.

$$Sensitivity = \frac{TP}{TP + FN}$$

3. Specificity:

$$Specificity = \frac{TN}{TN + FP}$$

Specificity (**true negative rate**) is calculated as the number of correct negative predictions divided by the total number of negatives.

TN= 150 ; FP = 28

4. Precision:

FN= 32 ; TP = 51

Precision (**Positive predictive value**) is calculated as the number of correct positive predictions divided by the total number of positive predictions.

$$Precision = \frac{TP}{TP + FP}$$

5. KS statistic : KS statistic is a measure of degree of separation between the positive and negative distributions. KS value of 100 indicates that the scores partition the records exactly such that one group contains all positives and the other contains all negatives. In practical situations, a KS value higher than 50% is desirable.

6. ROC chart & Area under the curve (AUC)

ROC chart is a plot of 1-specificity in the X axis and sensitivity in the Y axis. Area under the ROC curve is a measure of model performance. The AUC of a random classifier is 50% and that of a perfect classifier is 100%.

For practical situations, an AUC of over 70% is desirable.

7. Precision vs. recall: Recall or sensitivity gives us information about a model's performance on false negatives (incorrect prediction of customers who will default),

while precision gives us information of the model's performance of false positives.

8. F-measure [measure of a test's accuracy]

= F1 Score = 2*(Recall * Precision) / (Recall + Precision)

TN= 150 ; FP = 28

FN= 32 ; TP = 51

(F1 score or F score): alternate terms

Performance measures

- Performance metrics

$$Sensitivity\ (\%) = \frac{TP}{TP+FN}\ X\ 100$$

$$Specificity\ (\%) = \frac{TN}{TN+FP}\ X\ 100$$

$$Precision\ (\%) = \frac{TP}{TP+FP}\ X100$$

$$Recall\ (\%) = Sensistivity\ (\%) = \frac{TP}{TP+FN}\ X\ 100$$

$$Accuracy\ (\%) = \frac{TP+TN}{TP+FP+TN+FN}\ X\ 100$$

Performance measures Summary

		Actual class		
		Positive	Negative	
Predicted class	Positive	**TP:** True Positive	**FP:** False Positive (Type I Error)	**Precision:** $\frac{TP}{(TP + FP)}$
	Negative	**FN:** False Negative (Type II Error)	**TN:** True Negative	**Negative Predictive Value:** $\frac{TN}{(TN+FN)}$
		Recall or Sensitivity: $\frac{TP}{(TP + FN)}$	**Specificity:** $\frac{TN}{(TN + FP)}$	**Accuracy:** $\frac{TP + TN}{(TP + TN + FP + FN)}$

Rules extracted from classification algorithms

S.No	Rule
1	(Coughing = Yes) => Class=Pneumonia (352.0/0.0)
2	(Abnormal_Breathing = Yes) and (Depression = No) => Class=Pneumonia (74.0/0.0)
3	(Temprature = Normal) and (Fever = Yes) and (Decrease_Milk_yield = No) and (Nodular_Lesion = No) and (Painful = No) and (Loss_of_Apptetite = Yes) and (Lameness = No) and (Depression = No) => Class=Pneumonia (7.0/1.0)
4	(Smaking_Lips = Yes) => Class=FMD (385.0/2.0)
5	(Sore_feet = Yes) => Class=FMD (34.0/0.0)
6	(Temprature = High) and (Painful = No) and (Loss_of_Apptetite = No) and (Swelling_Udder = No) and (Crepitation_Sound = No) and (Fever = Yes) and (Depression = No) => Class=FMD (28.0/3.0)
7	(Lameness = Yes) and (Crepitation_Sound = No) and (Painful = No) => Class=FMD (4.0/1.0)
8	(Swelling_Udder = Yes) and (Nodular_Lesion = No) => Class=Mastitis (400.0/0.0)
9	(Change_color_of_Milk = Yes) => Class=Mastitis (33.0/0.0)
10	(Temprature = High) and (Fever = No) and (Crepitation_Sound = No) and (Painful = No) => Class=Mastitis (14.0/0.0)

The final classification Rules are the actual classifier model use for prediction.

Challenge of Evaluation Metrics

1) Evaluation measures play a crucial role in both assessing the classification performance and guiding the classifier modeling.
2) In fact, the use of common metrics in imbalanced domains can lead to sub-optimal classification models and might produce misleading conclusions since these measures are insensitive to skewed domains.

✓ skewness is a measure of the asymmetry of the probability distribution.

ACTIVITY-11

Explore a classification problem case by considering any real-world domain application, formulate a confusion matrix through scenario assumption for the classifier model and investigate the various parameters to measure the performance of the classifier model.

Extra Query

1. Investigate the four test options to perform training and testing for Data analytic algorithms based on the dataset. What do the four test options mean and when do you use them?
2. Investigate the major challenges in context to the performance measures of the classifier models connecting to the real-world application scenario.

Cheers For the Great Patience!

Query Please?