STAT 215A Fall 2017 Week 10

Rebecca Barter

10/27/2017

Evaluating classification algorithms

There are many ways to evaluate the accuracy of a classification algorithm

Evaluating classification algorithms

There are many ways to evaluate the accuracy of a classification algorithm

- Test/validation set
- Cross-validation
- Confusion matrix
- TP, FP, TN, FN
- ROC curve
- Confidence interval for error

Confusion matrix

Hypothesized class \ True class	p	n
Y	True Positives	False Positives
N	False Negatives	True Negatives
Column totals:	P	N

$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

$$\text{precision} = \frac{TP}{TP+FP} \qquad \text{recall} = \frac{TP}{P}$$

$$\text{accuracy} = \frac{TP+TN}{P+N}$$

$$\text{F-measure} = \frac{2}{1/\text{precision}+1/\text{recall}}$$

Fig. 1. Confusion matrix and common performance metrics calculated from it.

Ref: Fawcett (2005)

Confusion matrix

		True class p	True class n
Hypothesized class	Y	True Positives	False Positives
	N	False Negatives	True Negatives
Column totals:		P	N

ROC curve

$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

$$\text{precision} = \frac{TP}{TP+FP} \qquad \text{recall} = \frac{TP}{P}$$

$$\text{accuracy} = \frac{TP+TN}{P+N}$$

$$\text{F-measure} = \frac{2}{1/\text{precision}+1/\text{recall}}$$

Fig. 1. Confusion matrix and common performance metrics calculated from it.

Ref: Fawcett (2005)

Confusion matrix

		True class	
		p	**n**
Hypothesized class	**Y**	True Positives	False Positives
	N	False Negatives	True Negatives
Column totals:		P	N

$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

precision-recall curve

$$\boxed{\text{precision} = \frac{TP}{TP+FP} \qquad \text{recall} = \frac{TP}{P}}$$

$$\text{accuracy} = \frac{TP+TN}{P+N}$$

$$\text{F-measure} = \frac{2}{1/\text{precision}+1/\text{recall}}$$

Fig. 1. Confusion matrix and common performance metrics calculated from it.

Ref: Fawcett (2005)

Receiver operating characteristics (ROC) graphs

Used to depict the tradeoff between “hit rates” and “false alarm rates” of classifiers.

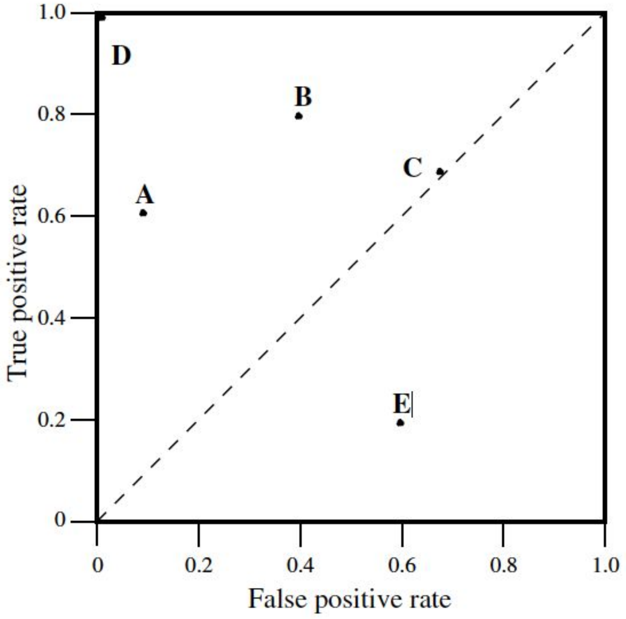


Fig. 2. A basic ROC graph showing five discrete classifiers.

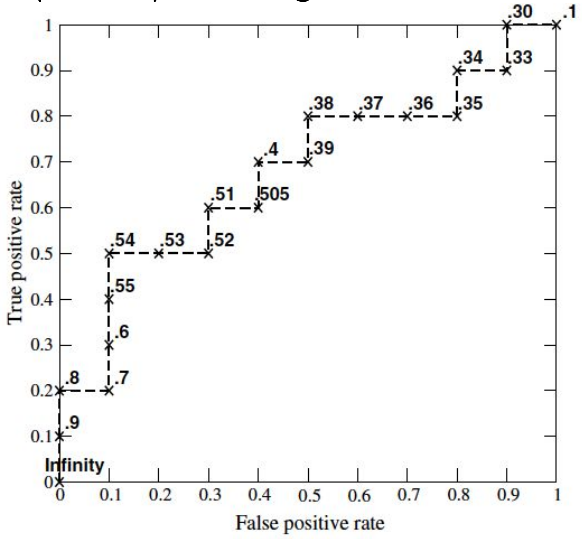
Receiver operating characteristics (ROC) graphs

When can we generate an ROC curve?

- When the output of a classifier is a probability and we must choose a threshold for the final predicted class.

Inst#	Class	Score	Inst#	Class	Score
1	p	.9	11	p	.4
2	p	.8	12	n	.39
3	n	.7	13	p	.38
4	p	.6	14	n	.37
5	p	.55	15	n	.36
6	p	.54	16	n	.35
7	n	.53	17	p	.34
8	n	.52	18	n	.33
9	p	.51	19	p	.30
10	n	.505	20	n	.1

(TP, FP) for the given threshold



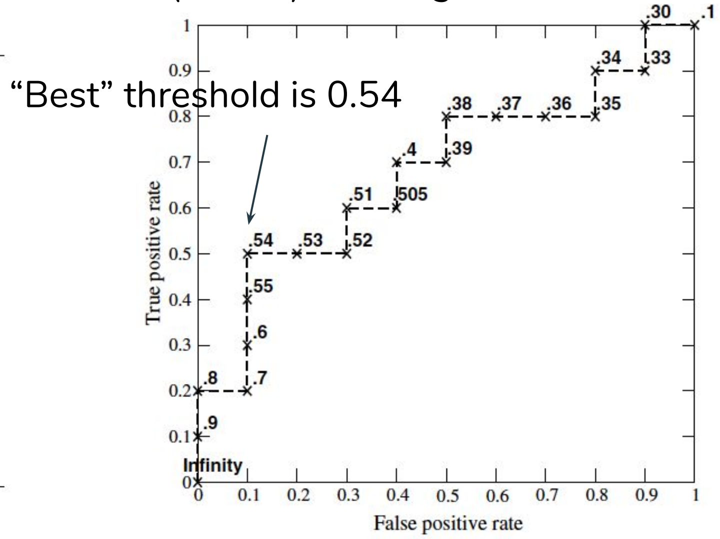
Receiver operating characteristics (ROC) graphs

When can we generate an ROC curve?

- When the output of a classifier is a probability and we must choose a threshold for the final predicted class.

Inst#	Class	Score	Inst#	Class	Score
1	p	.9	11	p	.4
2	p	.8	12	n	.39
3	n	.7	13	p	.38
4	p	.6	14	n	.37
5	p	.55	15	n	.36
6	p	.54	16	n	.35
7	n	.53	17	p	.34
8	n	.52	18	n	.33
9	p	.51	19	p	.30
10	n	.505	20	n	.1

(TP, FP) for the given threshold



Receiver operating characteristics (ROC) graphs

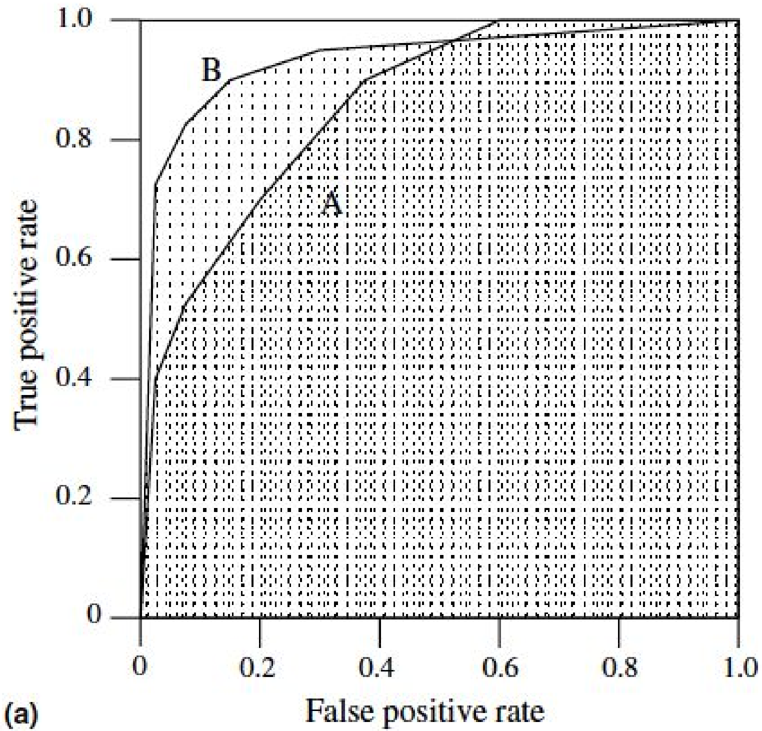
Questions:

Would this method give an accurate reflection when there is class imbalance?

(e.g. 10% of the observations are in group 1 and 90% are in group 2)

Receiver operating characteristics (ROC) graphs

The area under the curve (AUC) is a method for comparing algorithms



(a)

The Area Under The Curve (AUC)

The area under the curve (AUC) is a method for evaluating classifier accuracy.

The AUC has an important statistical property:

- *the AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance*

The Area Under The Curve (AUC)

Care should be taken when using ROC curves to compare classifiers

- *The ROC graph is often used to select the best classifiers simply by graphing them in ROC space and seeing which ones dominate.*

The Area Under The Curve (AUC)

Care should be taken when using ROC curves to compare classifiers

- *The ROC graph is often used to select the best classifiers simply by graphing them in ROC space and seeing which ones dominate.*
- *This is misleading; it is analogous to taking the maximum of a set of accuracy figures from a single test set.*

The Area Under The Curve (AUC)

Care should be taken when using ROC curves to compare classifiers

- *The ROC graph is often used to select the best classifiers simply by graphing them in ROC space and seeing which ones dominate.*
- *This is misleading; it is analogous to taking the maximum of a set of accuracy figures from a single test set.*
- *Without a measure of variance we cannot compare the classifiers.*

The Area Under The Curve (AUC)

Care should be taken when using ROC curves to compare classifiers

- *The ROC graph is often used to select the best classifiers simply by graphing them in ROC space and seeing which ones dominate.*
- *This is misleading; it is analogous to taking the maximum of a set of accuracy figures from a single test set.*
- *Without a measure of <u>variance</u> we cannot compare the classifiers.*

It is a good idea to take the ***average*** *of multiple ROC curves (e.g. via cross validation)*

See Fawcett (2005) for examples on how to average.

The Area Under The Curve (AUC)

Example: classify.R

Lab 4: cloud detection

Groups

lab4_groups.csv

name	group
Hongxu Ma	1
Alexander Brandt	1
Jordan Prosky	1
Chandan Singh	2
Ningning Long	2
Zoe Vernon	2
Nan Yang	3
Rosanna Neuhausler	3
Runjing Liu	3
Junyu Cao	4
Xu Rao	4
Max Gardner	4
Alan Dong	5
Xiao Li	5
Amy Ko	5
Donghyeon Ko	6
Eric Kim	6
Ruonan Hao	6
Yizhou Zhao	7
Miyabi Ishihara	7
Olivia Angiuli	7
Zhiyi You	8
DA XU	8
Orhan Ocal	8
Xiaoqi Zhang	9
Aniket Kesari	9
Mingjia Chen	9
Meng QI	10
Feynman Liang	10
Weicheng Kuo	10
Shuhui Huang	11
Lily Xue Gong	11
Jake Soloff	11
Hector Roux de Bezieux	12
Briton Park	12

Data: satellite images over the arctic region

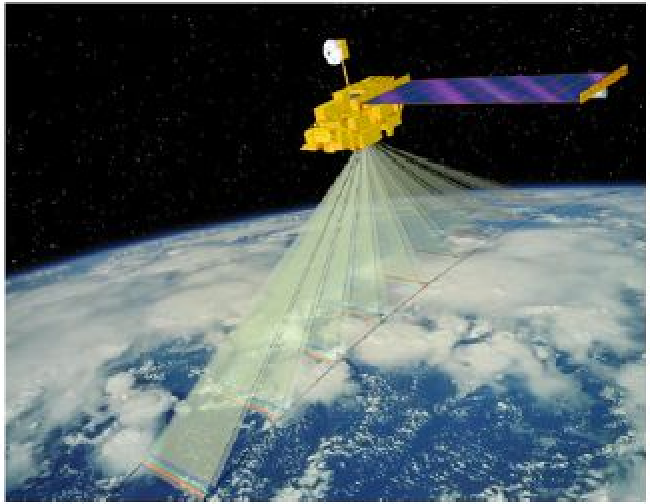


Figure 1. Cartoon illustration of the Terra satellite with the view directions of the nine MISR cameras. Image is courtesy of the MISR science team at the Jet Propulsion Laboratory.

Data: satellite images over the arctic region

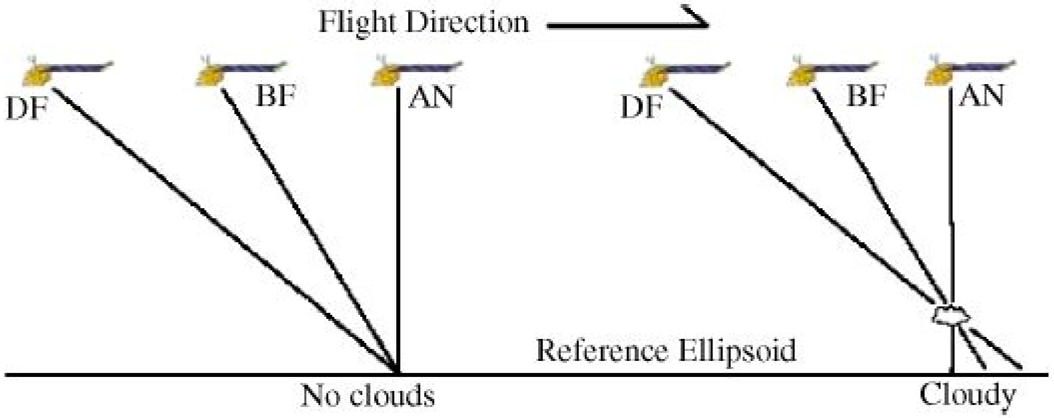


Figure 2. Registration of surface features and clouds to the reference ellipsoid. Note that only three of the nine MISR cameras are illustrated and that surface objects are registered to the same location, whereas clouds are registered to different locations.

Data: satellite images over the arctic region

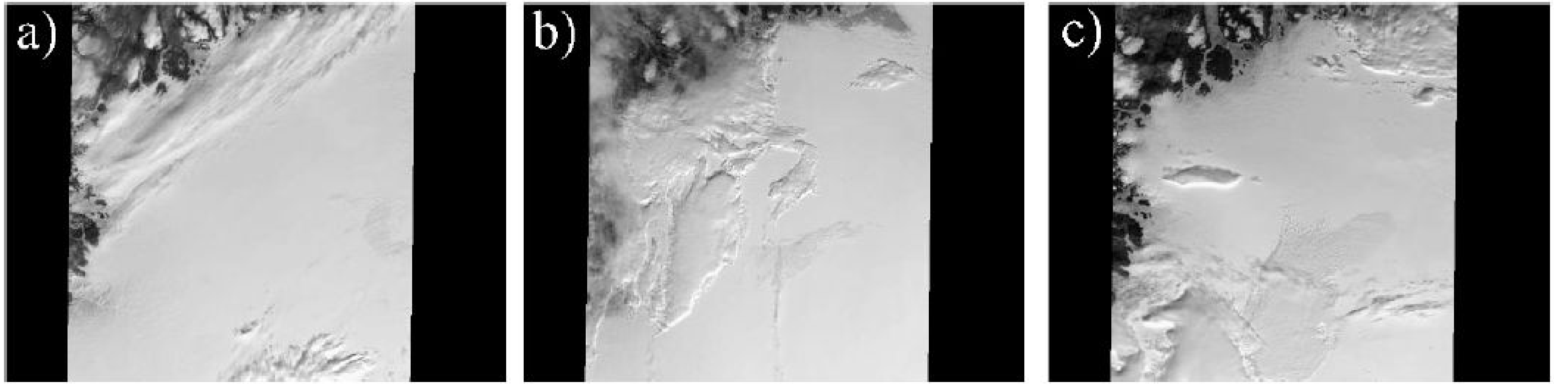


Figure 3. Data collected by the MISR An-camera for three consecutive orbits (i.e., 13257, 13490, and 13723) over blocks 20–22 of path 26.

Data: satellite images over the arctic region

Figure 4. Expert labels for blocks 20–22 of MISR orbits (a) 13257, (b) 13490, and (c) 13723. White represents high confidence cloudy; gray, high confidence clear; and black, unlabeled pixels.

The Area Under The Curve (AUC)

Example: cloud_data.R