

Model Evaluation and Improvement

Dr. Yuzana Win (Nagasaki University, Japan)
Lecturer
Department of Computer Engineering and
Information Technology

Lecture Objectives

- To introduce
 - Why would we need evaluation?
 - Model Evaluation Method

Why would we need evaluation?

- Establish model programs and best practices by providing feedback about what we worked and what we failed
- Measures impact the program is making how to improve the performance



Model Evaluation

- Metrics for Performance Evaluation
 - How to evaluate the performance of a model?
- Methods for Performance Evaluation
 - How to obtain reliable estimates?

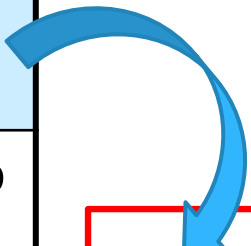
Model Evaluation

- **Metrics for Performance Evaluation**
 - How to evaluate the performance of a model?
- **Methods for Performance Evaluation**
 - How to obtain reliable estimates?

Metrics for Performance Evaluation

- Focus on the predictive capability of a model, using **Confusion Matrix**

		PREDICTED CLASS	
		Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	a	b
	Class=No	c	d



a: TP (true positive)
b: FN (false negative)
c: FP (false positive)
d: TN (true negative)

Metrics for Performance Evaluation

		PREDICTED CLASS	
		Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	a (TP)	b (FN)
	Class=No	c (FP)	d (TN)

$$\text{Accuracy} = \frac{a + d}{a + b + c + d} = \frac{TP + TN}{TP + TN + FP + FN}$$

Cost Matrix

	PREDICTED CLASS		
ACTUAL CLASS	$C(i j)$	Class=Yes	Class=No
	Class=Yes	$C(\text{Yes} \text{Yes})$	$C(\text{No} \text{Yes})$
	Class=No	$C(\text{Yes} \text{No})$	$C(\text{No} \text{No})$

$C(i|j)$: Cost of misclassifying class j example as class i

Computing Cost of Classification

Cost Matrix	PREDICTED CLASS		
	C(i j)	+	-
	+	-1	100
	-	1	0

Model M₁	PREDICTED CLASS		
		+	-
	+	150	40
	-	60	250

Accuracy = 80%

Cost = 3910

Model M₂	PREDICTED CLASS		
		+	-
	+	250	45
	-	5	200

Accuracy = 90%

Cost = 4255

Cost vs Accuracy

Count	PREDICTED CLASS		
	Class=Yes	Class=No	
ACTUAL CLASS	Class=Yes	a	b
	Class=No	c	d

Cost	PREDICTED CLASS		
	Class=Yes	Class=No	
ACTUAL CLASS	Class=Yes	p	q
	Class=No	q	p

Accuracy is proportional to cost if

1. $C(\text{Yes}|\text{No})=C(\text{No}|\text{Yes}) = q$
2. $C(\text{Yes}|\text{Yes})=C(\text{No}|\text{No}) = p$

$$N = a + b + c + d$$

$$\text{Accuracy} = (a + d)/N$$

$$\begin{aligned} \text{Cost} &= p(a + d) + q(b + c) \\ &= p(a + d) + q(N - a - d) \\ &= qN - (q - p)(a + d) \\ &= N[q - (q - p) \times \text{Accuracy}] \end{aligned}$$

Cost-Sensitive Measures

$$\text{Precision (p)} = \frac{a}{a + c}$$

$$\text{Recall (r)} = \frac{a}{a + b}$$

$$\text{F - measure (F)} = \frac{2rp}{r + p} = \frac{2a}{2a + b + c}$$

- Precision is biased towards $C(\text{Yes}|\text{Yes})$ & $C(\text{Yes}|\text{No})$
- Recall is biased towards $C(\text{Yes}|\text{Yes})$ & $C(\text{No}|\text{Yes})$
- F-measure is biased towards all except $C(\text{No}|\text{No})$

Model Evaluation

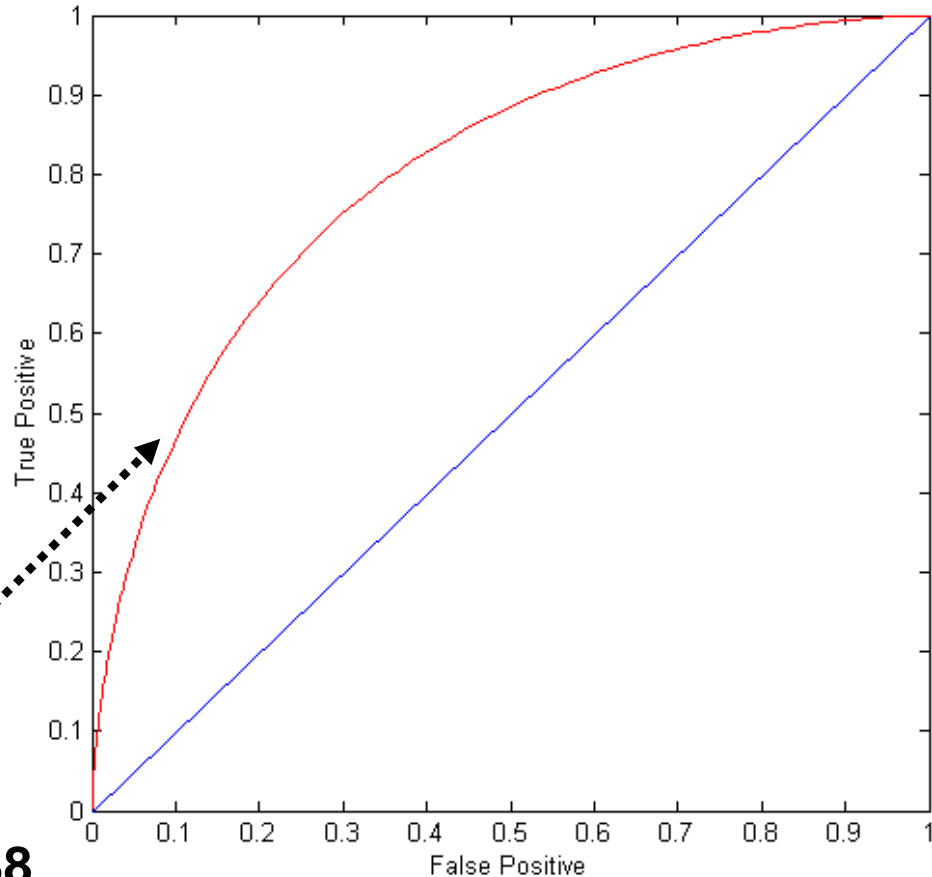
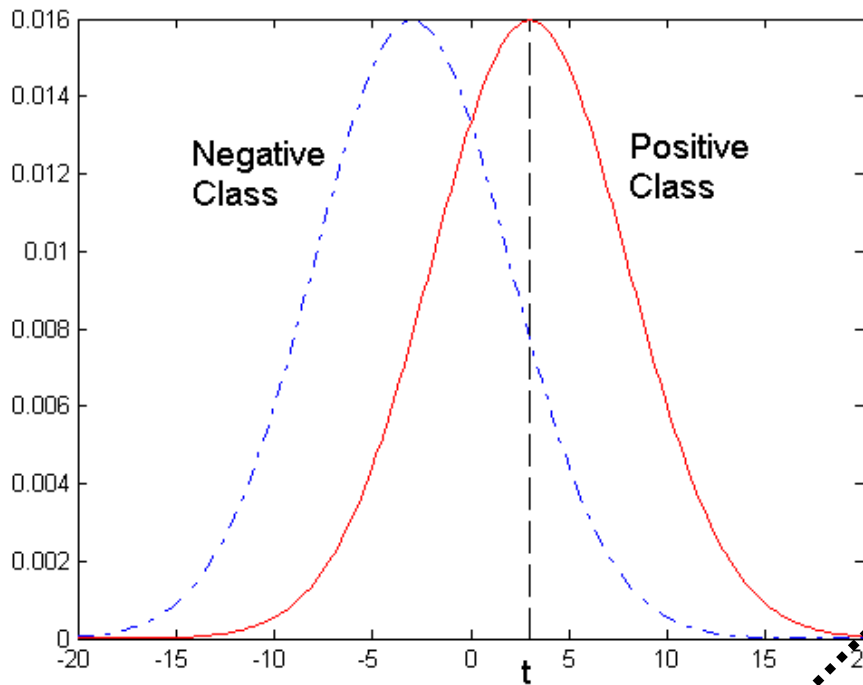
- Metrics for Performance Evaluation
 - How to evaluate the performance of a model?
- **Methods for Performance Evaluation**
 - How to obtain reliable estimates?

Methods for Performance Evaluation

- How to obtain a reliable estimate of performance?
 - **Holdout**
 - Reserve $2/3$ for training and $1/3$ for testing
 - **Random subsampling**
 - Repeated holdout
 - **Cross validation**
 - Partition data into k disjoint subsets
 - k -fold: train on $k-1$ partitions, test on the remaining one
 - Leave-one-out: $k=n$
 - **Stratified sampling**
 - oversampling vs undersampling
 - **Bootstrap**
 - Sampling with replacement

ROC (Receiver Operating Characteristic)

- 1-dimensional data set containing 2 classes (positive and negative)
- any points located at $x > t$ is classified as positive



At threshold t :

TP=0.5, FN=0.5, FP=0.12, FN=0.88

How to Construct an ROC curve

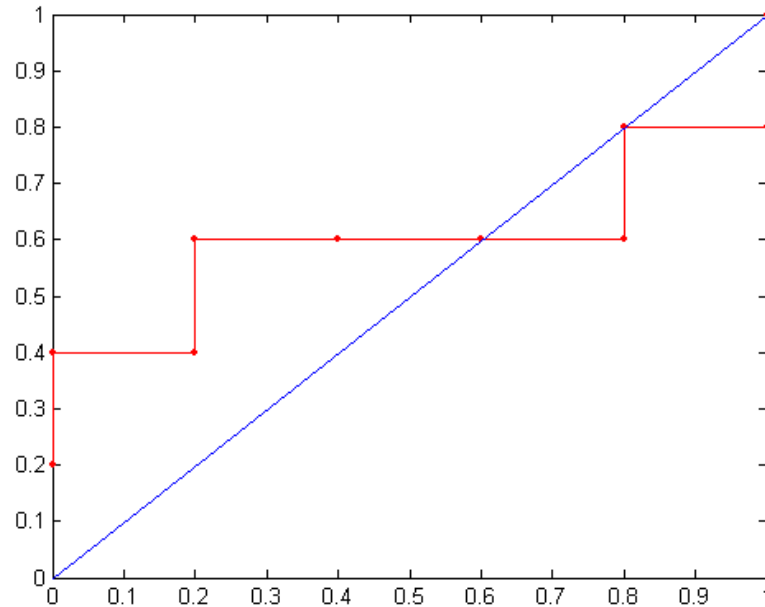
Instance	$P(+ A)$	True Class
1	0.95	+
2	0.93	+
3	0.87	-
4	0.85	-
5	0.85	-
6	0.85	+
7	0.76	-
8	0.53	+
9	0.43	-
10	0.25	+

- Use classifier that produces posterior probability for each test instance $P(+|A)$
- Sort the instances according to $P(+|A)$ in decreasing order
- Apply threshold at each unique value of $P(+|A)$
- Count the number of TP, FP, TN, FN at each threshold
- TP rate, $TPR = TP/(TP+FN)$
- FP rate, $FPR = FP/(FP + TN)$

How to Construct an ROC curve

Class	+	-	+	-	-	-	+	-	+	+	
Threshold \geq	0.25	0.43	0.53	0.76	0.85	0.85	0.85	0.87	0.93	0.95	1.00
TP	5	4	4	3	3	3	3	2	2	1	0
FP	5	5	4	4	3	2	1	1	0	0	0
TN	0	0	1	1	2	3	4	4	5	5	5
FN	0	1	1	2	2	2	2	3	3	4	5
TPR	1	0.8	0.8	0.6	0.6	0.6	0.6	0.4	0.4	0.2	0
FPR	1	1	0.8	0.8	0.6	0.4	0.2	0.2	0	0	0

ROC Curve:



References

- Pang-Ning Tan, Michael Steinbach, Vipin Kumar: *Introduction to Data Mining*, Addison-Wesley

Next Week Lecture

- Practical: Working with Text Data