Model Selection and Evaluation

Jin Guo

Activity1

Model Selection and Evaluation

1. Identify the metrics you know
2. Any observations form this table

	Accuracy	Root mean Square error	True positive Rate	False Positive Rate	Precision	Recall	F-measure		Kononenko and Bratko's information score
Algorithm	Acc	RMSE	TPR	FPR	Prec	Rec	*F*	AUC	K & B
NB	71.7	0.4534	0.44	0.16	0.53	0.44	0.48	0.7	48.1118
C45	75.5	0.4324	0.27	0.04	0.74	0.27	0.4	0.59	34.2789
3NN	72.4	0.5101	0.32	0.1	0.56	0.32	0.41	0.63	43.3682
RIP	71	0.4494	0.37	0.14	0.52	0.37	0.43	0.6	22.3397
SVM	69.6	0.5515	0.33	0.15	0.48	0.33	0.39	0.59	54.8934
Bagging	67.8	0.4518	0.17	0.1	0.4	0.17	0.23	0.63	11.3004
Boosting	70.3	0.4329	0.42	0.18	0.5	0.42	0.46	0.7	34.4795
RF	69.23	0.47	0.33	0.15	0.48	0.33	0.39	0.63	20.7763

Area under the ROC curve

Observations

Model Selection and Evaluation

Algorithm	Acc	RMSE	TPR	FPR	Prec	Rec	*F*	AUC	K & B
NB	71.7	0.4534	0.44	0.16	0.53	0.44	0.48	0.7	48.1118
C45	75.5	0.4324	0.27	0.04	0.74	0.27	0.4	0.59	34.2789
3NN	72.4	0.5101	0.32	0.1	0.56	0.32	0.41	0.63	43.3682
RIP	71	0.4494	0.37	0.14	0.52	0.37	0.43	0.6	22.3397
SVM	69.6	0.5515	0.33	0.15	0.48	0.33	0.39	0.59	54.8934
Bagging	67.8	0.4518	0.17	0.1	0.4	0.17	0.23	0.63	11.3004
Boosting	70.3	0.4329	0.42	0.18	0.5	0.42	0.46	0.7	34.4795
RF	69.23	0.47	0.33	0.15	0.48	0.33	0.39	0.63	20.7763

Accuracy

Binary classification case

Accuracy

Input	Actual Output	Model Output
1	*No*	*Yes*
2	*No*	*No*
3	*Yes*	*No*
.....	*.....*	*.....*

Accuracy

Binary classification case

	Model Pred Positive	Model Pred Negative
Actual Positive	*True Positive(TP)*	*False Negative(FN)*
Actual Negative	*False Positive(FP)*	*True Negative(TN)*

Accuracy *(TP + TN) / (P + N)*

Input	Actual Output	Model Output
1	*No*	*Yes*
2	*No*	*No*
3	*Yes*	*No*
.....	*.....*	*.....*

Activity 2

Binary classification case

- Think of two applications in which one area in the confusion matrix is more important than others.

	Model Pred Positive	Model Pred Negative
Actual Positive	*True Positive(TP)*	*False Negative(FN)*
Actual Negative	*False Positive(FP)*	*True Negative(TN)*

Input	Actual Output	Model Output
1	*No*	*Yes*
2	*No*	*No*
3	*Yes*	*No*
……	*……*	*……*

Confusion Matrix

Binary classification case

Precision	*TP / (TP + FP)*
Recall	*TP / (TP + FN)*
F - measure	$\frac{(1+\alpha)Precision * Recall}{\alpha * Precision + Recall}$

	Model Pred Positive	**Model Pred Negative**
Actual Positive	*True Positive(TP)*	*False Negative(FN)*
Actual Negative	*False Positive(FP)*	*True Negative(TN)*

	Model Pred Positive	Model Pred Negative
Actual Positive	100	300
Actual Negative	200	200

Confusion Matrix

Binary classification case

	Model Pred Positive	Model Pred Negative
Actual Positive	*True Positive(TP)*	*False Negative(FN)*
Actual Negative	*False Positive(FP)*	*True Negative(TN)*

Precision *TP / (TP + FP)* 100/(100+300) = 0.25

Recall *TP / (TP + FN)* 100/(100+200) = 0.33

F - measure

$$\frac{(1+\alpha)Precision * Recall}{\alpha * Precision + Recall}$$

F1 = 0.28
F2 = 0.3
F0.5 = 0.27

	Model Pred Positive	Model Pred Negative
Actual Positive	100	300
Actual Negative	200	200

Confusion Matrix

Binary classification case

	Model Pred Positive	Model Pred Negative
Actual Positive	*True Positive(TP)*	*False Negative(FN)*
Actual Negative	*False Positive(FP)*	*True Negative(TN)*

Precision *TP / (TP + FP)* 100/(100+300) = 0.25

Recall *TP / (TP + FN)* 100/(100+200) = 0.33

F - measure $\frac{(1+\alpha)Precision * Recall}{\alpha * Precision + Recall}$

F1 = 0.28
F2 = 0.3
F0.5 = 0.27

	Model Pred Positive	Model Pred Negative
Actual Positive	100	300
Actual Negative	200	200 -> 0

False Positive Rate (fallout) *FP / (FP + TN)*

True Positive Rate (Hit Rate) *TP / (TP + FN)*

Confusion Matrix

Binary classification case

	Model Pred Positive	**Model Pred Negative**
Actual Positive	*True Positive(TP)*	*False Negative(FN)*
Actual Negative	*False Positive(FP)*	*True Negative(TN)*

There is no single metric is capable of encapsulating all aspect of interest!

Scoring Classifier?

Binary classification case

Impact of the threshold?

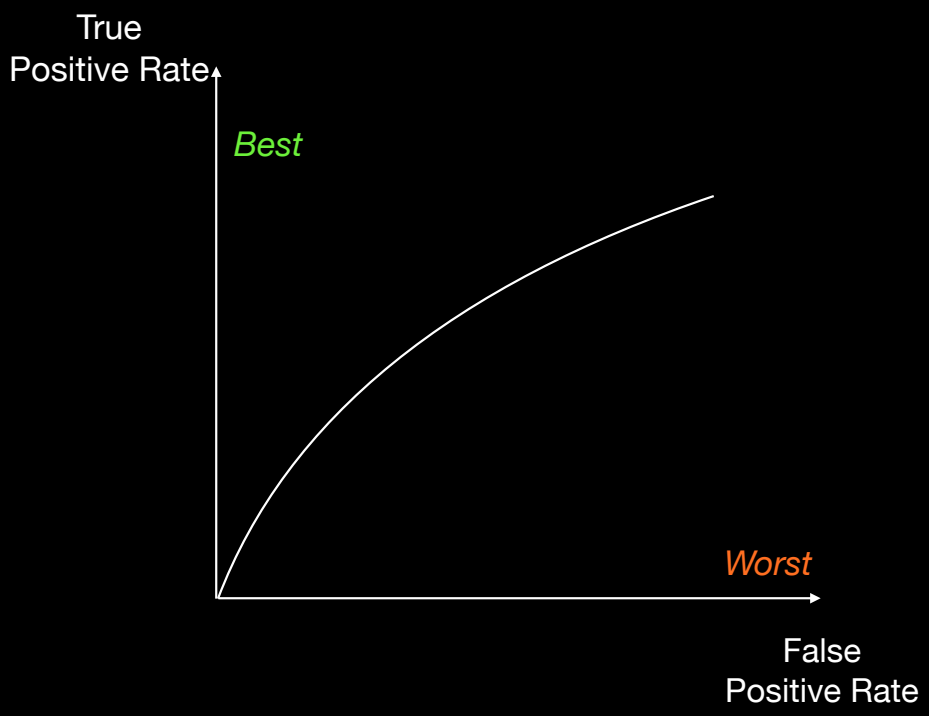
	Model Pred Positive	Model Pred Negative
Actual Positive	*True Positive(TP)*	*False Negative(FN)*
Actual Negative	*False Positive(FP)*	*True Negative(TN)*

Input	Actual Output	Model Output
1	*No*	*0.8 -> Yes*
2	*No*	*0.7 -> No*
3	*Yes*	*0.75 -> No*
.....	*.....*	*.....*
14	*Yes*	*0.4*

Precision-Recall (PR) Curves

Input	Actual Output	Model Output
1	No	0.8 -> Yes
2	No	0.7 -> No
3	Yes	0.75 -> No
.....	.....	.....
14	Yes	0.4

ROC Analysis



	Model Pred Positive	Model Pred Negative
Actual Positive	*True Positive(TP)*	*False Negative(FN)*
Actual Negative	*False Positive(FP)*	*True Negative(TN)*

True Positive Rate (Hit Rate) *TP / (TP + FN)*

False Positive Rate (fallout) *FP / (FP + TN)*

Metrics discussed so far

- Confusion matrix based
- Considering the class distribution and classifier uncertainty

Domain specific metrics

- Mean Average precision
  - Information Retrieval
- BLEU
  - Text generation
- ...

How to choose a baseline?

- Random Classifier?
- Problem domain?
- State of the art?

Algorithm	Acc	RMSE	TPR	FPR	Prec	Rec	*F*	AUC	K & B
NB	71.7	0.4534	0.44	0.16	0.53	0.44	0.48	0.7	48.1118
C45	75.5	0.4324	0.27	0.04	0.74	0.27	0.4	0.59	34.2789
3NN	72.4	0.5101	0.32	0.1	0.56	0.32	0.41	0.63	43.3682
RIP	71	0.4494	0.37	0.14	0.52	0.37	0.43	0.6	22.3397
SVM	69.6	0.5515	0.33	0.15	0.48	0.33	0.39	0.59	54.8934
Bagging	67.8	0.4518	0.17	0.1	0.4	0.17	0.23	0.63	11.3004
Boosting	70.3	0.4329	0.42	0.18	0.5	0.42	0.46	0.7	34.4795
RF	69.23	0.47	0.33	0.15	0.48	0.33	0.39	0.63	20.7763