Metrics of Multi-labels Classification

©Kuang

Graduate School of Information, Production and Systems
早稲田大学　大学院情報生産システム研究科

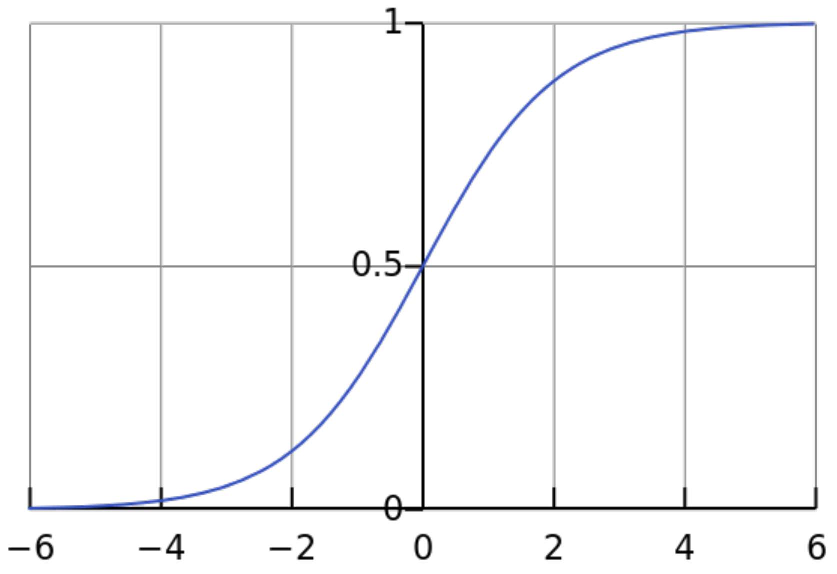
Background

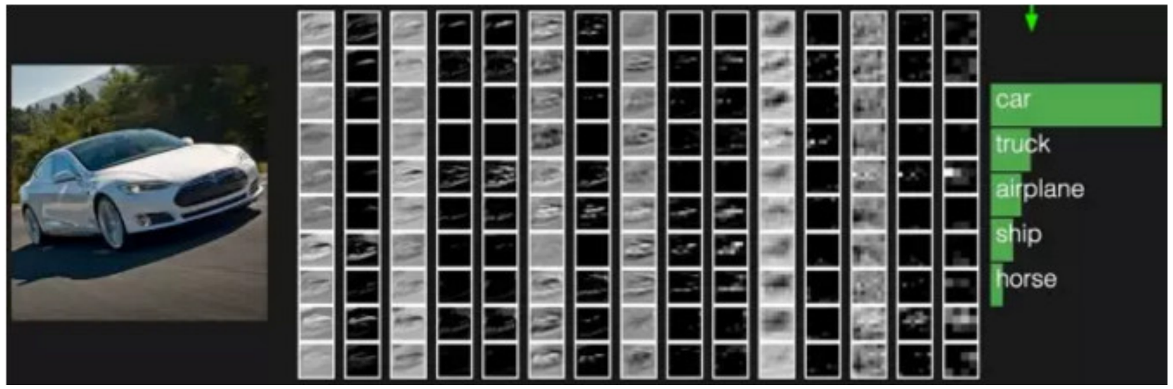
Binary Classification

A single instance (feature vector) associated with a single label.

Multi-Class Learning

A single instance (feature vector) which its classes are mutually exclusive and each sample can't belong to several classes simultaneously..





[1]Zhang M L, Zhou Z H. A review on multi-label learning algorithms[J]. IEEE transactions on knowledge and data engineering, 2014, 26(8): 1819-1837.

[2] Chen B, Gu W, Hu J. An improved multi-label classification based on label ranking and delicate boundary SVM[C]//Neural Networks (IJCNN), The 2010 International Joint Conference on. IEEE, 2010: 1-6.

Background

“Multi-label learning studies the problem where each example is represented by a single instance while associated with a set of labels simultaneously.”[1]

[1]Chen B, Gu W, Hu J. An improved multi-label classification based on label ranking and delicate boundary SVM[C]//Neural Networks (IJCNN), The 2010 International Joint Conference on. IEEE, 2010: 1-6.

Background

Shôshanku no sora ni (1994)

Awards

Showing all 19 wins and 37 nominations

Academy Awards, USA 1995

Nominated Oscar	Best Picture Niki Marvin
	Best Actor in a Leading Role Morgan Freeman
	Best Writing, Screenplay Based on Material Previously Produced or Published Frank Darabont
	Best Cinematography Roger Deakins
	Best Sound Robert J. Litt Elliot Tyson Michael Herbick Willie D. Burton
	Best Film Editing Richard Francis-Bruce
	Best Music, Original Score Thomas Newman

Background

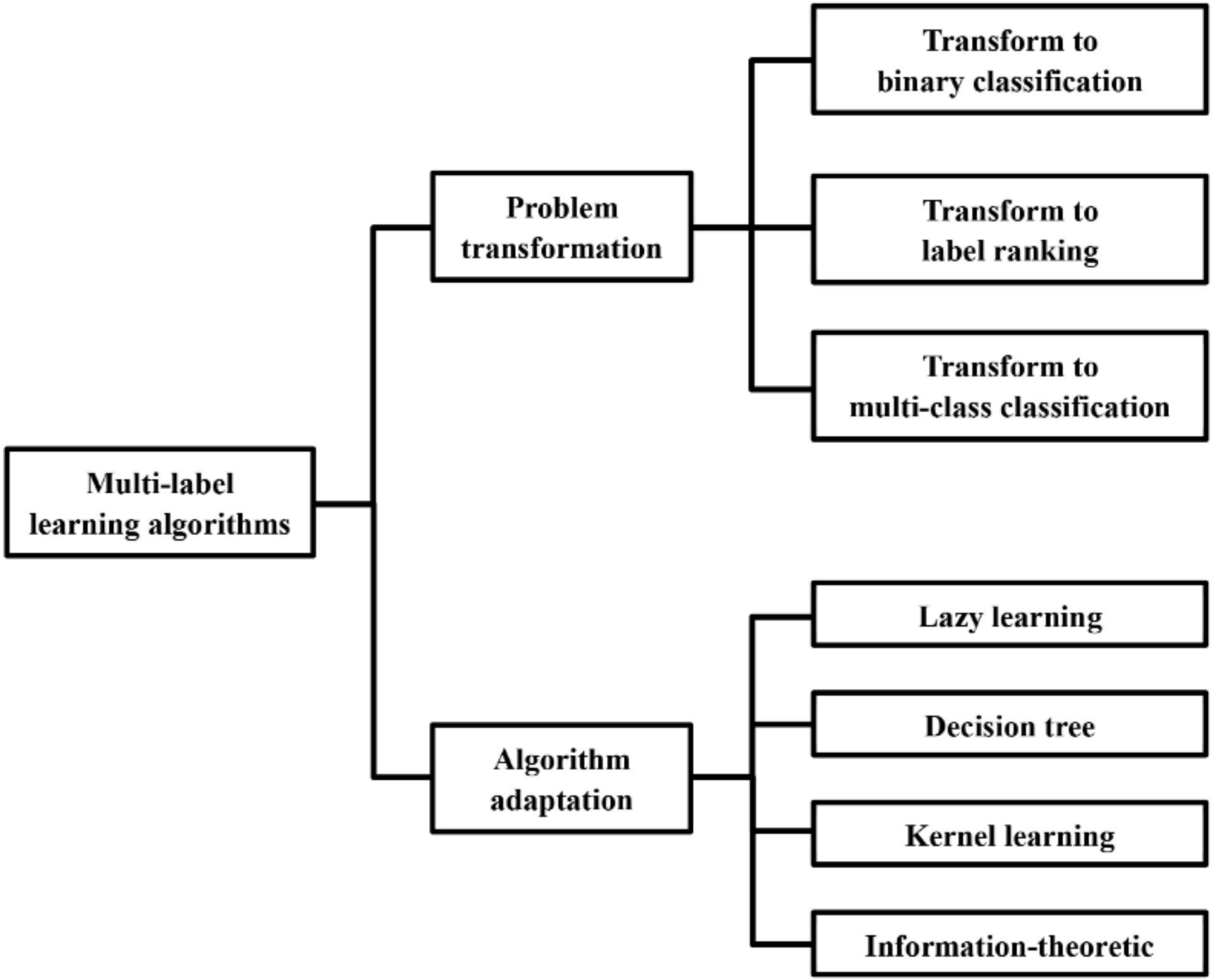
Key Challenge

The key challenge of learning from multi-label data lies in the overwhelming size of output space, the number of label sets grows exponentially as the number of class labels increases.

a label space with 20 class labels

the number of possible label sets would be 2^20

Categorization of Multi-label Learning



Background

Binary Relevance

In one-versus-rest methods, the multi-label training set is simply divided into m (the number of labels) binary class subsets.

Multi-labels Learning

A single instance associated with m labels.

A single instance associated with 1 labels.

A single instance associated with 1 labels.

…

A single instance associated with 1 labels.

m

Chen B, Gu W, Hu J. An improved multi-label classification based on label ranking and delicate boundary SVM[C]//Neural Networks (IJCNN), The 2010 International Joint Conference on. IEEE, 2010: 1-6.

Metrics of Binary Classification

Focus on: Accuracy, Precision, Recall and F Score

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

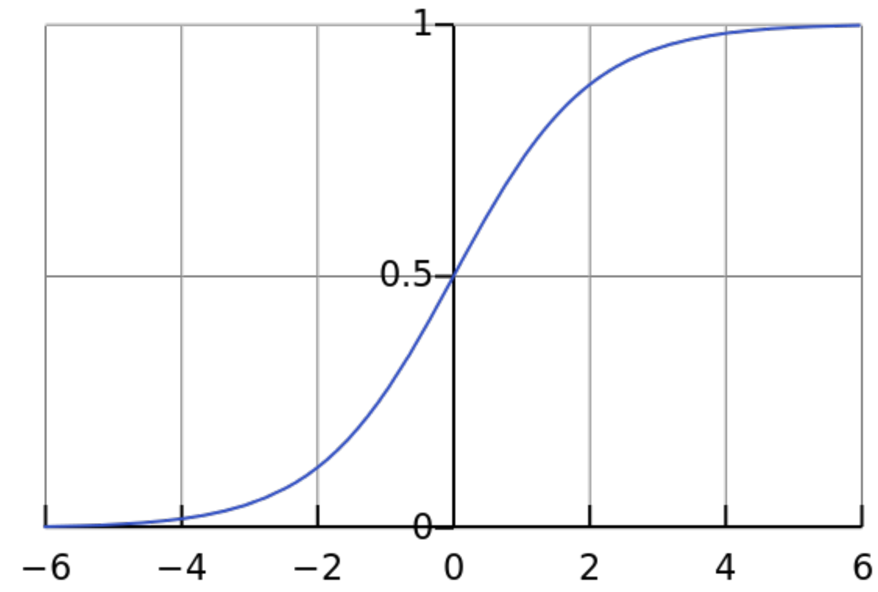
$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$$

tp: True Positive
tn: True Negative
fp: False Positive
fn: False Negative



Sigmoid Function

周志华，机器学习

Metrics of Binary Classification

Accuracy

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

Assume, there is a data set with 99% negative samples and 1% positive Samples.

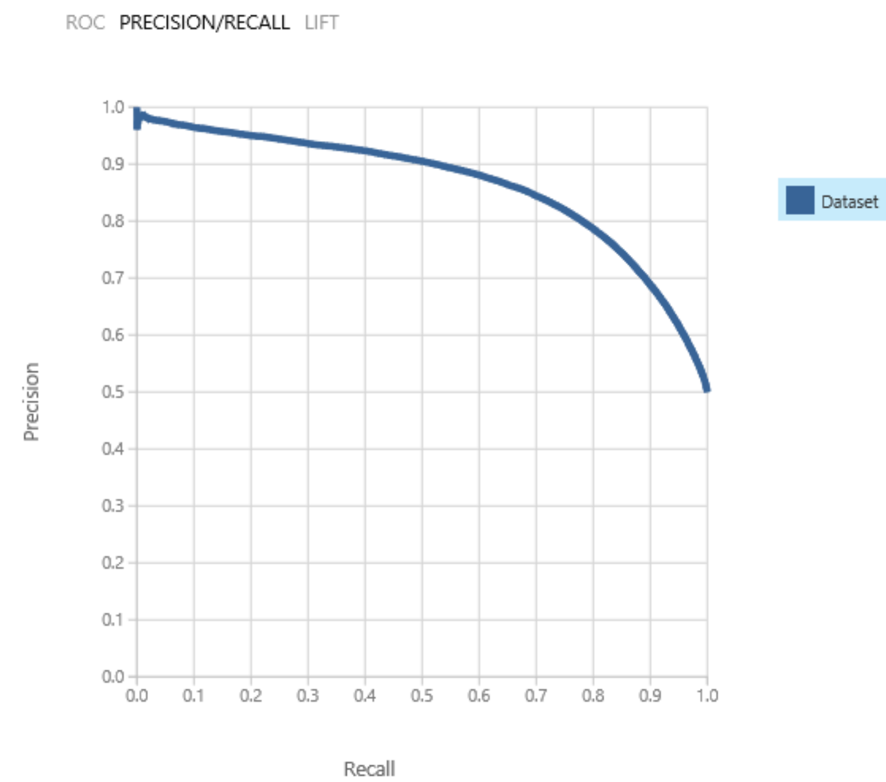
A nonsense classifier just output negative, can get a high accuracy like 99%.

Metrics of Binary Classification

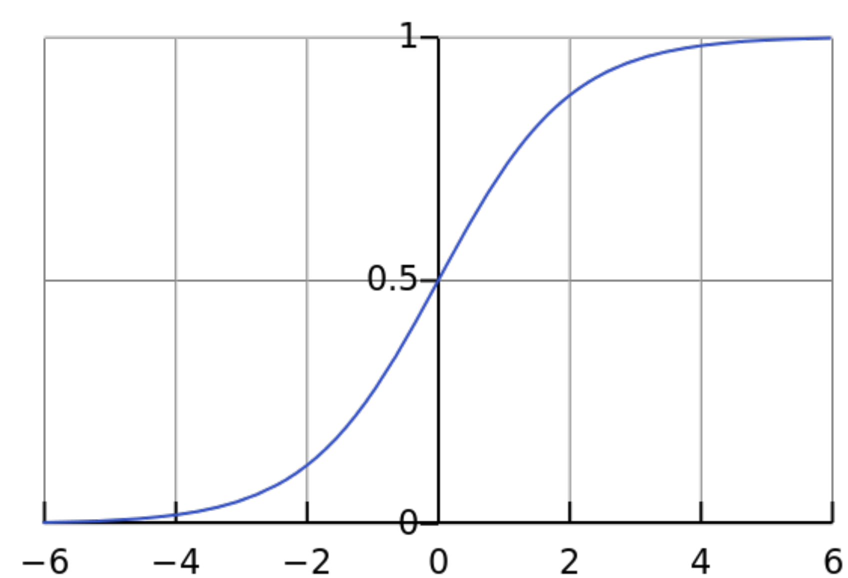
Precision and Recall

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
129808	30192	0.790	0.778	0.5	0.865
False Positive	True Negative	Recall	F1 Score		
37074	122926	0.811	0.794		

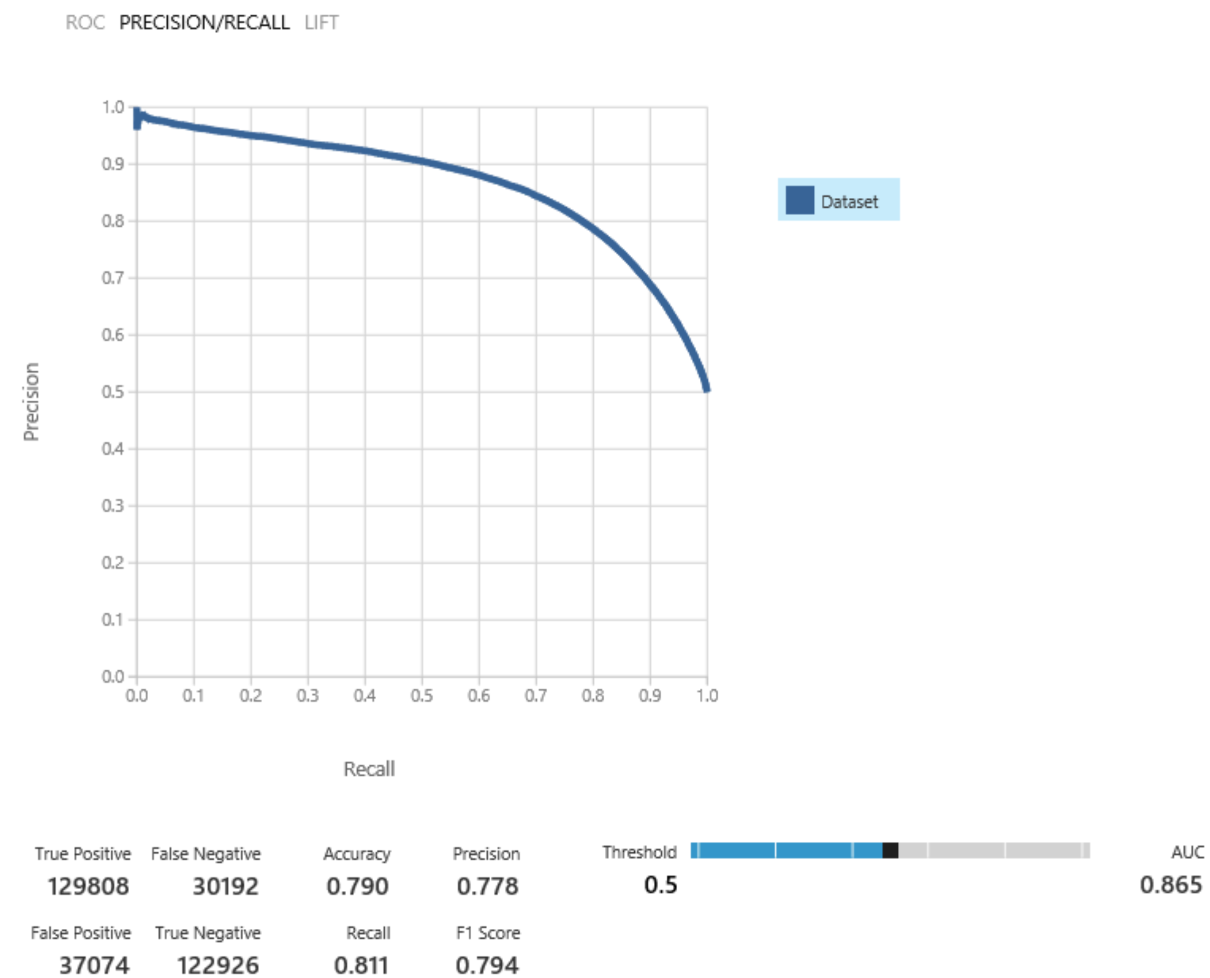


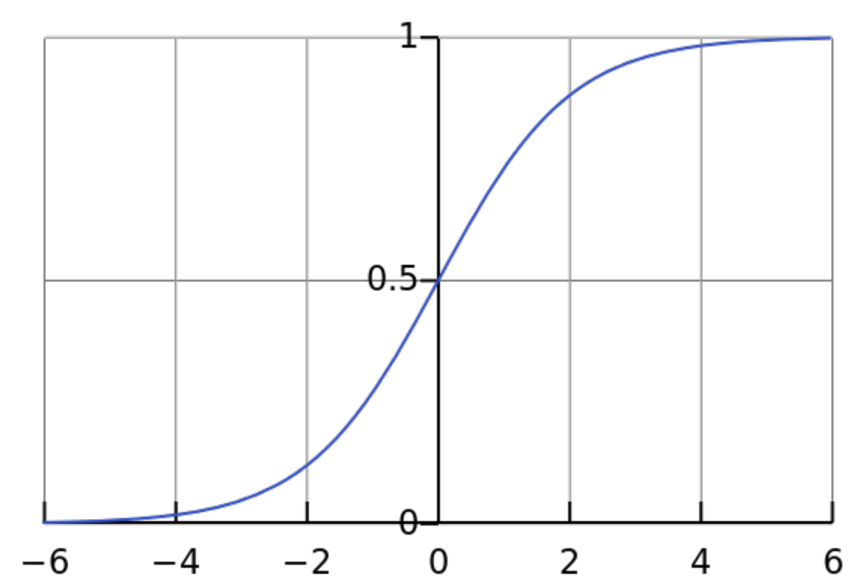
Metrics of Binary Classification

Precision and Recall

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

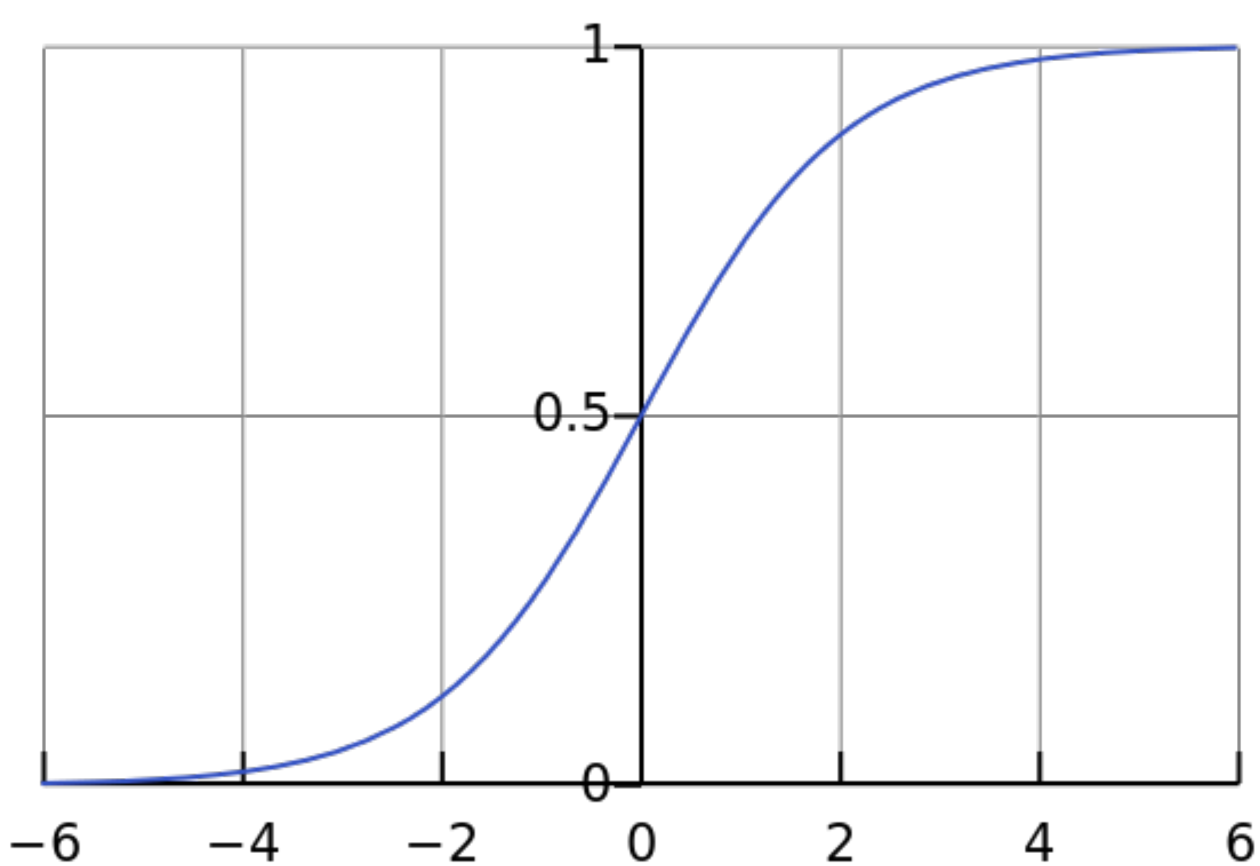


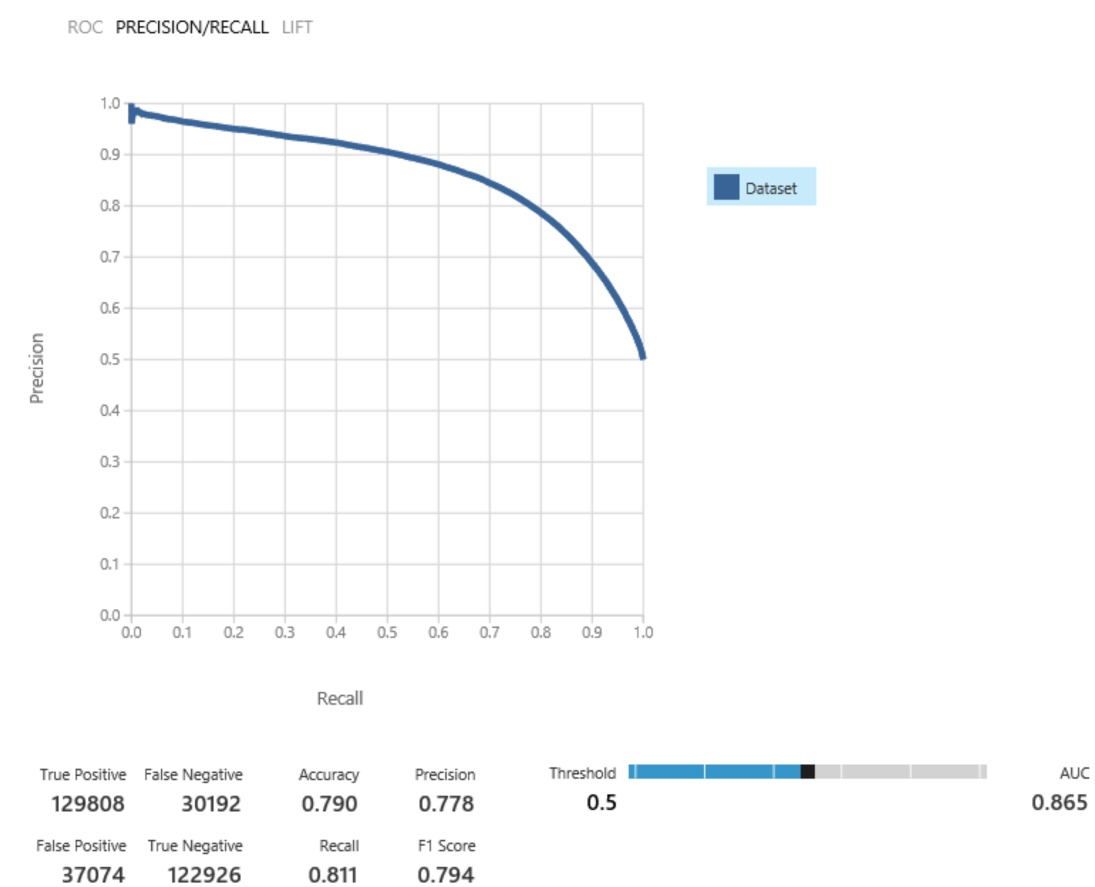


Metrics of Binary Classification

Threshold and Ranking

Most of binary classifier calculate probability of prediction, and the classifier compare the probability with the threshold.





Metrics of Binary Classification

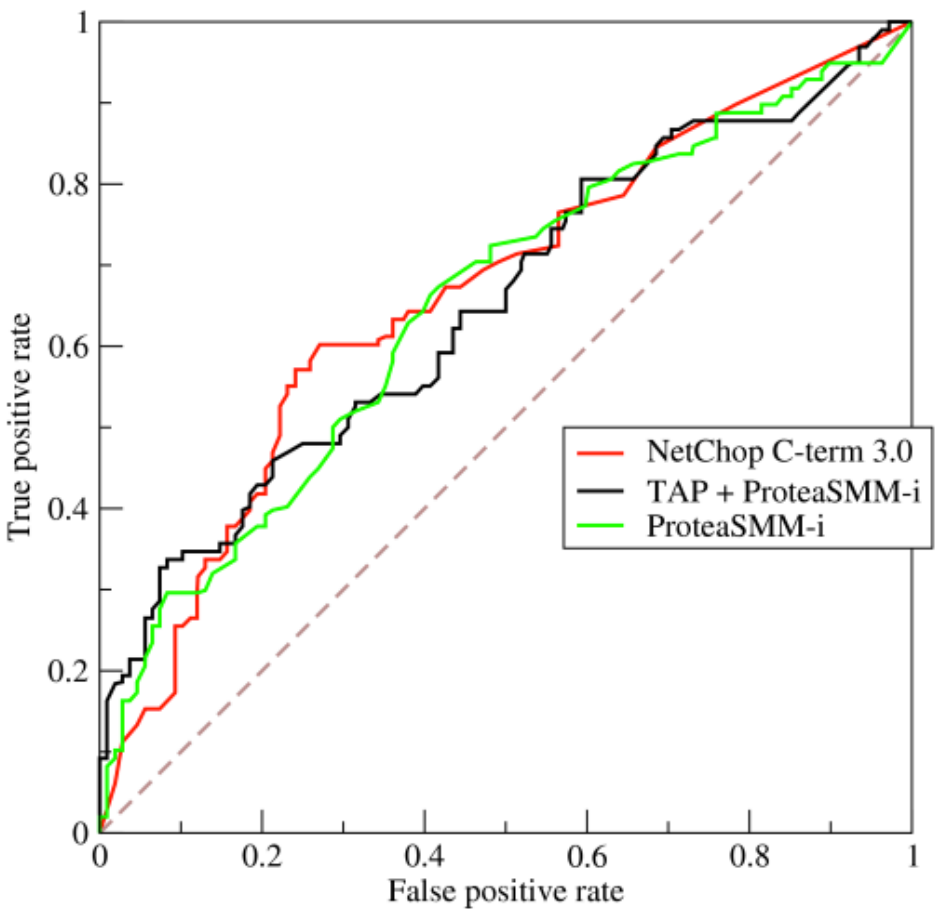
ROC and AUC

The ROC curve is created by plotting the **true positive rate** (TPR) against the **false positive rate** (FPR) at various threshold settings.

$$\mathrm{TPR} = \frac{\mathrm{TP}}{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{TNR} = \frac{\mathrm{TN}}{N} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

* AUC,Area Under the ROC Curve, a parameter to evaluate classifier.



周志华，机器学习

Metrics of Binary Classification

F Score

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Equal Cost, β==1

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$$

Unequal Cost, β>1 **Recall** get higher weight, β>1 **Precision** get higher weight

周志华，机器学习

Metrics of Binary Classification

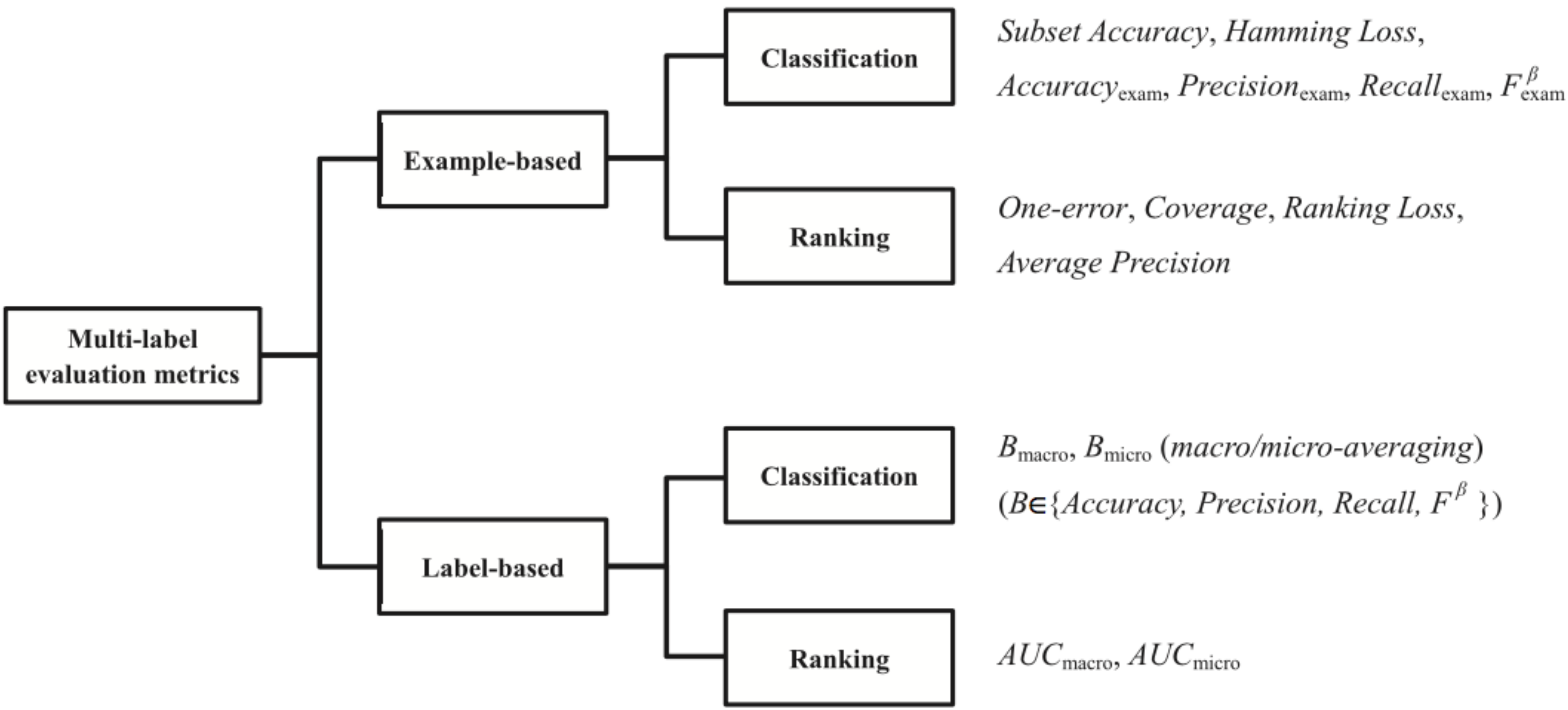
	GO:0016020	GO:0005634	GO:0016021	GO:0008150	GO:0009507	GO:0046872	GO:0005886	GO:0005737	GO:0006355	GO:00167
Accuracy	0.927298	0.933888	0.960586	0.951081	0.938746	0.978371	0.930213	0.93655	0.988679	0.974527
Precision	0.951925	0.913446	0.949326	0.802923	0.763702	0.927426	0.80335	0.836555	0.947623	0.875981
Recall	0.783547	0.76247	0.864486	0.697094	0.597976	0.933005	0.535856	0.627554	0.952577	0.93062
F1-Score	0.859568	0.831158	0.904922	0.746275	0.670754	0.930207	0.642888	0.717137	0.950093	0.902475

...	GO:0009570	GO:0003735	GO:0003676	GO:0006412	GO:0006508	GO:0005774	GO:0055085	GO:0005622	GO:0005618	GO:0005575
...	0.969458	0.976681	0.965106	0.971274	0.970767	0.97119	0.968951	0.967979	0.971401	0.985891
...	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
...	0	0	0	0	0	0	0	0	0	0
...	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

: The larger the metric value, the better algorithm performance

Metrics of Multi-Label Classification

Example-based metrics work by evaluating the learning system's performance on each test example separately, and then returning the mean value across the test set.
Label-based metrics work by evaluating the learning system's performance on each class label separately, and then returning the macro/micro-averaged value across all class labels.



Metrics of Multi-Label Classification

Example-based metrics work by evaluating the learning system's performance on each test example separately, and then returning the mean value across the test set.

Focus on: Subset Accuracy, Hamming Loss…

$$hloss(h) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{q}\,|h(\boldsymbol{x}_i)\Delta Y_i|$$

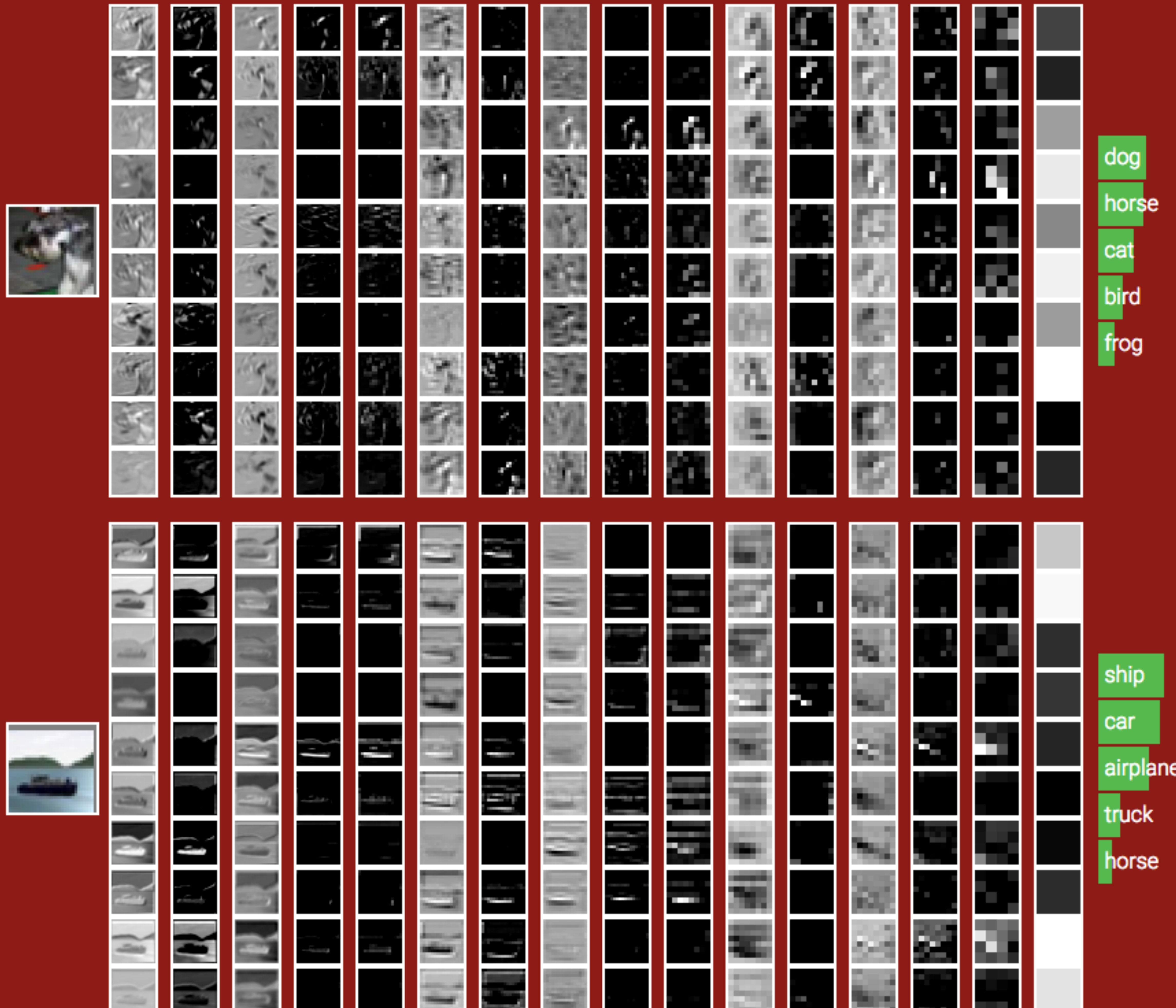
$$subsetacc(h) = \frac{1}{p}\sum_{i=1}^{p}[\![h(\boldsymbol{x}_i) = Y_i]\!]$$

Metrics of Multi-Label Classification

$$Precision_{\text{exam}}(h) = \frac{1}{p}\sum_{i=1}^{p}\frac{|Y_i \bigcap h(\boldsymbol{x}_i)|}{|h(\boldsymbol{x}_i)|}$$

$$Recall_{\text{exam}}(h) = \frac{1}{p}\sum_{i=1}^{p}\frac{|Y_i \bigcap h(\boldsymbol{x}_i)|}{|Y_i|}$$

$$F^{\beta}_{\text{exam}}(h) = \frac{(1+\beta^2)\cdot Precision_{\text{exam}}(h)\cdot Recall_{\text{exam}}(h)}{\beta^2\cdot Precision_{\text{exam}}(h) + Recall_{\text{exam}}(h)}$$


dog
horse
cat
bird
frog
ship
car
airplane
truck
horse

Metrics of Multi-Label Classification

$$one\text{-}error(f) = \frac{1}{p}\sum_{i=1}^{p} [\![[\arg\max_{y\in\mathcal{Y}} f(\boldsymbol{x}_i, y)] \notin Y_i]\!].$$

$$coverage(f) = \frac{1}{p}\sum_{i=1}^{p} \max_{y\in Y_i} rank_f(\boldsymbol{x}_i, y) - 1.$$

$$rloss(f) = \frac{1}{p}\sum_{i=1}^{p} \frac{1}{|Y_i||\bar{Y}_i|} |\{(y', y'') \mid f(\boldsymbol{x}_i, y') \leq f(\boldsymbol{x}_i, y''),\ (y', y'') \in Y_i \times \bar{Y}_i)\}|.$$

Metrics of Multi-Label Classification

Label-based metrics work by evaluating the learning system's performance on each class label separately, and then returning the macro/micro-averaged value across all class labels.

$$B_{\mathrm{macro}}(h) = \frac{1}{q}\sum_{j=1}^{q} B(TP_j, FP_j, TN_j, FN_j)$$

$$B_{\mathrm{micro}}(h) = B\left(\sum_{j=1}^{q} TP_j,\ \sum_{j=1}^{q} FP_j,\ \sum_{j=1}^{q} TN_j,\ \sum_{j=1}^{q} FN_j\right)$$

$$per-classrecall = \frac{1}{C}\sum_{i=1}^{c}\frac{N_i^c}{N_i^g}$$

$$per-classprecision = \frac{1}{C}\sum_{i=1}^{c}\frac{N_i^c}{N_i^p}$$

$$overall-recall = \frac{\sum_{i=1}^{c} N_i^c}{\sum_{i=1}^{c} N_i^g}$$

$$overall-precision = \frac{\sum_{i=1}^{c} N_i^c}{\sum_{i=1}^{c} N_i^p}$$

Metrics of Multi-Label Classification

Macro-Precision:	0.874
Macro-Recall:	0.743
Macro-F1 Score:	0.803
Macro-Accuracy :	0.969

: The larger the metric value, the better algorithm performance