Area Under the Curve

Lesson 7 – Section 2

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Overview

- Review of recall, precision, Type I error, and Type II error
- ROC Curve and Area Under Curve (AUC)
- Statistical Meaning of AUC

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Performance Metrics in Classification

		Actual	
		Positive	Negative
Predicted	Pos	n_{11}	n_{12}
	Neg	n_{21}	n_{22}

- Type I error (False positive):

$$\Pr(\hat{y} = Pos \mid y = Neg) = n_{12}/(n_{12} + n_{22}) = n_{12}/n_{\bullet 2}$$

- Type II error(False negative):

$$\Pr(\hat{y} = Neg \mid y = Pos) = n_{21}/(n_{11} + n_{21}) = n_{21}/n_{\bullet 1}$$

- Accuracy: $(n_{11} + n_{22})/(n_{11} + n_{22} + n_{12} + n_{21})$
- Sensitivity (True Positive Rate, Recall): Among the (n_{11}+n_{21}) true positive cases, the percentage that is predicted as positive:

$$\Pr(\hat{y} = Pos \mid y = Pos) = n_{11}/(n_{11} + n_{21}) = n_{11}/n_{\bullet 1} = 1 - Type\ II\ \text{Error}$$

- Precision: Among the (n_{11}+n_{12}) predicted positive cases, the percentage that is actually positive:

$$\Pr(y = Pos \mid \hat{y} = Pos) = n_{11}/(n_{11} + n_{12}) = n_{11}/n_{1\bullet}$$

ROC Curve

- Two performance metrics are used to plot ROC curve:
  - True Positive (1-Type II error): $\Pr(\hat{y} = Pos \mid y = Pos) = n_{11}/(n_{11} + n_{21}) = n_{11}/n_{\bullet 1} = 1 - Type\ II\ \text{Error}$
  - False Positive (Type I error): $\Pr(\hat{y} = Pos \mid y = Neg) = n_{12}/(n_{12} + n_{22}) = n_{12}/n_{\bullet 2}$
  - If threshold = 0, all cases are classified as Positive, Type I error = 1 (all negative cases are classified as Positive), True Positive Rate = 1
  - If threshold = 1, all cases are classified as Negative, Type I error = 0 (all negative cases are classified as Negative), True Positive Rate = 0 (All Positive cases are classified as Negative)
- By changing the threshold between 0 and 1, we get a curve connecting [0,0] and [1,1]
- This curve is called ROC curve. Receiver Operating Characteristic

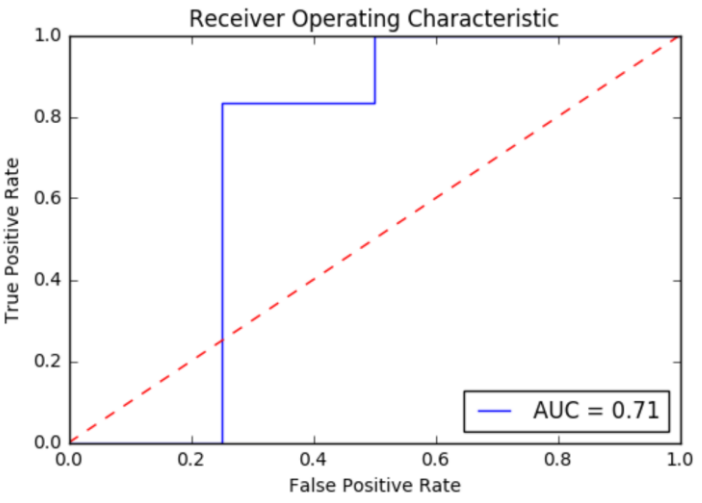
Example

Education	Age	Employer Sector	Prob(Salary>65k)	Actual	Threshold
High-School	50	Government	0.38	0	0.5
Bachelor	45	Private	0.69	1	
Associate	30	Non-profit	0.61	0	
Bachelor	36	Private	0.73	1	
Master	42	Private	0.82	1	
PhD	48	Government	0.7	1	
Master	25	Private	0.56	1	
Associate	20	Non-profit	0.48	0	
Bachelor	37	Private	0.92	0	
PhD	51	Government	0.79	1	

		Actual	
		Positive	Negative
Predict	Positive	6	2
	Negative	0	2

Recall=	100
Precision=	75

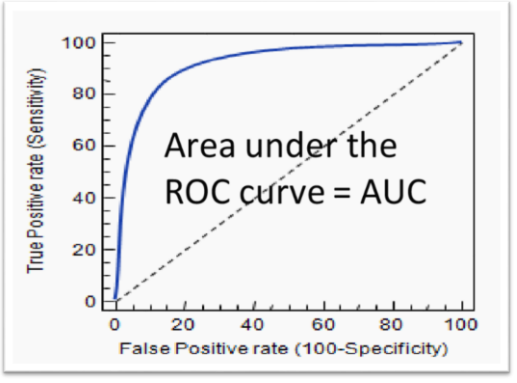
```
import matplotlib.pyplot as plt
plt.title('Receiver Operating Characteristic')
plt.plot(fpr, tpr, 'b', label = 'AUC = %0.2f' % auc)
plt.legend(loc = 'lower right')
plt.plot([0, 1], [0, 1],'r--')
plt.xlim([0, 1])
plt.ylim([0, 1])
plt.ylabel('True Positive Rate')
plt.xlabel('False Positive Rate')
plt.show()
```



ROC Curve

- Area under the ROC curve (AUC) is a measure of model performance.

  0.5 (*random model*) < ***AUC*** < 1 (*perfect model*)



- Larger the AUC, better the model

What Is the Statistical Meaning of AUC?

$$\Pr(\hat{y}_i > \hat{y}_j \mid y_i > y_j), \forall i \neq j$$

- Example:

```
import numpy as np
from sklearn import metrics

truth = np.array([0,1,0,1])
prediction = np.array([0.2, 0.5, 0.6, 0.45])

fpr, tpr, thresholds = metrics.roc_curve(truth, prediction, pos_label=1)
auc = metrics.auc(fpr, tpr)
print("Y=" + ",".join(map(str, truth)) + ", Y_hat=%s, AUC=%.2f"%(",".join(map(str, prediction)), auc))

prediction = np.array([0.2, 0.5, 0.6, 0.70])
fpr, tpr, thresholds = metrics.roc_curve(truth, prediction, pos_label=1)
auc = metrics.auc(fpr, tpr)
print("Y=" + ",".join(map(str, truth)) + ", Y_hat=%s, AUC=%.2f"%(",".join(map(str, prediction)), auc))
```

```
Y=0,1,0,1, Y_hat=0.2,0.5,0.6,0.45, AUC=0.50
Y=0,1,0,1, Y_hat=0.2,0.5,0.6,0.7, AUC=0.75
```

yi > yj	$\hat{y}_i > \hat{y}_j$
y2 > y1	Yes
y2 > y3	No
y4 > y1	Yes
y4 > y3	No
AUC	0.5

yi > yj	$\hat{y}_i > \hat{y}_j$
y2 > y1	Yes
y2 > y3	No
y4 > y1	Yes
y4 > y3	Yes
AUC	0.75

Change the last prediction to 0.90 does not change AUC

When to Use AUC as the Performance Metrics?

- When ranking matters
  - –Recommendation systems: positions of the returned items from search significantly impact the click through rate. So, we need to have the items that are most likely clicked by users returned in higher positions in the search results
  - –Marketing applications: You have 1M customers, but your marketing campaign only has budget to reach out to 1k customers. You need to rank your customers based on their predicted response probability.

When to Use AUC as the Performance Metrics?

- When training data is extremely skewed:
  - –0.5 is not a reasonable threshold any more.
  - –If it is not easy to choose a threshold such that the cost of Type I and Type II errors can be minimized, start with AUC as your performance metrics

However

- Most of the machine learning models are implemented to minimize RMSE-like loss function
  - –One model with smaller RMSE does not ensure it will have higher AUC
  - –You should tune model parameters to optimize AUC
- Some machine learning algorithms are implemented to optimize AUC
  - –Choose these algorithms if you really want to optimize AUC during the model training process
  - –SVM^{perf} provides option to optimize AUC during model training process https://www.cs.cornell.edu/people/tj/svm_light/svm_perf.html

Performance Metrics

Receiver operating characteristic (ROC) curve, AUC metric…

- 1.0: perfect prediction *!Alert: Too Good To Be True*
- 0.9: excellent prediction
- 0.8: good prediction
- 0.7: mediocre prediction
- 0.6: poor prediction
- 0.5: random prediction
- <0.5: something wrong!

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Summary

>ROC Curve
- Receiver operating characteristic
- TPR vs FPR at various threshold settings

>AUC and its statistical meaning
- Area under the ROC curve

>When to choose AUC
- Ranking
- Skewed training data