Learning Fair Scoring Functions
Bipartite Ranking under ROC-based Fairness Constraints

Louise Davy

Télécom Paris

May 30, 2024

Introduction
oo

Definitions
ooooooooooooooooo

Contributions
ooooooooo

Applications
ooooooo

Conclusion
ooo

Outline

1 Introduction

2 Definitions
   Ranking, bipartite ranking and pairwise bipartite ranking
   ROC curve and AUC
   ROC and bipartite ranking

3 Contributions
   AUC-based constraints
   ROC based constraints

4 Applications
   Compas dataset
   Adult dataset
   Results

5 Conclusion

Outline

1 Introduction

2 Definitions

3 Contributions

4 Applications

5 Conclusion

Introduction

Paper

- **Title** : Learning Fair Scoring Functions : Bipartite Ranking under ROC-based Fairness Constraints

- **Authors** : Robin Vogel, Aurélien Bellet, Stephan Clémençon

- **Year** : 2021

- **Arxiv link** : https://arxiv.org/abs/2002.08159

- **Github link** : https://github.com/RobinVogel/Learning-Fair-Scoring-Functions

- **Blogpost link** :
  https://responsible-ai-datascience-ipparis.github.io/posts/lambert-davy/

Outline

Outline

Ranking

What is ranking ?
Ranking is a class of machine learning algorithms aiming to
sort a list of observations according to some **criterion**.

Examples

- Information retrieval : Sort documents according to their
  relevance to a query
- Recommendation systems : Recommend user's favourite
  songs first

Bipartite ranking

What is bipartite ranking ?
In bipartite ranking, we consider that all the observations that
we want to sort can be partitioned into two classes : **positive**
and **negative**. We want the positive instances to be consistently
ranked higher than the negative ones.

Examples

- Fraud detection : Find the observations that are most
  likely to be fraudulent among fraudulent and
  non-fraudulent observations

- Recommendation systems : Recommend user's favourite
  songs first but this time we have songs that are liked by the
  user and songs that are disliked
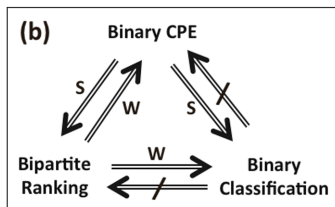
Bipartite ranking

What is the difference between bipartite ranking and binary classification ?
Bipartite ranking is very close to binary classification since we are trying to distinguish positive instances from negative instances, but **serves a slightly different goal**. In the cases where a model needs to process a **large number of observations** and where a **human verification is needed**, a bipartite ranking model would be able to provide the **most likely positive instances** first, allowing the human to only investigate a limited number of instances.

Introduction
oo

Definitions
ooooo●oooooooooooo

Contributions
oooooooo

Applications
ooooooo

Conclusion
ooo

Bipartite ranking

What is the difference between bipartite ranking and binary classification ?

There are some works[1] working around the link between the two that were able to show that a good ranking model, once transferred to binary classification, will perform well (provided that the right threshold was found), while the opposite is not always true.

---

[1] Narasimhan and Agarwal, "On the relationship between binary classification, bipartite ranking, and binary class probability estimation".

Pairwise bipartite ranking

What is pairwise bipartite ranking ?
Pairwise bipartite ranking is specific case of bipartite ranking, in which we rank each instance **relatively to another instance**. Instead of simply distinguishing between positive and negative items, pairwise bipartite ranking considers the **relative preference between pairs of items**.

Example
- Facial recognition : Find pairs of faces that are the most similar in a database

(This is not the focus of this presentation, but this is what I'm currently working on.)

Outline

Introduction
○○

Definitions
○○○○○○○○○●○○○○○○○○○

Contributions
○○○○○○○○

Applications
○○○○○○○

Conclusion
○○○

ROC curve

What is a ROC curve ?

ROC stands for **Receiver Operating Characteristic** curve and is a graph showing the performance of a classification model **at all classification thresholds**. It plots the false positive rate in the x-axis against the true positive rate in the y-axis.
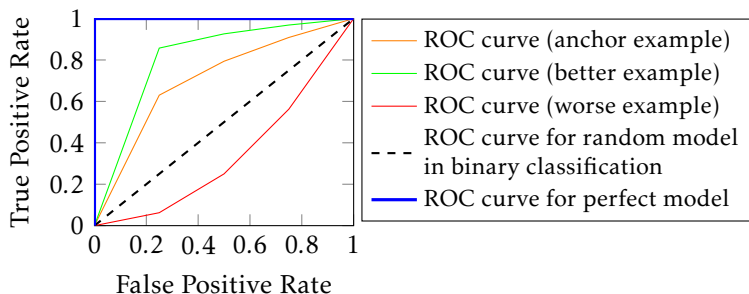

Figure: Different ROC curves

AUC

What is the AUC ?
AUC stands for **Area Under the Curve** and is a widely used
metric for machine learning model evaluation that quantifies
the overall performance of the model **across all possible
classification thresholds**. AUC measures the entire
two-dimensional area underneath the entire ROC curve from
(0,0) to (1,1).
Example

- A model who is 100% wrong has an AUC of 0.
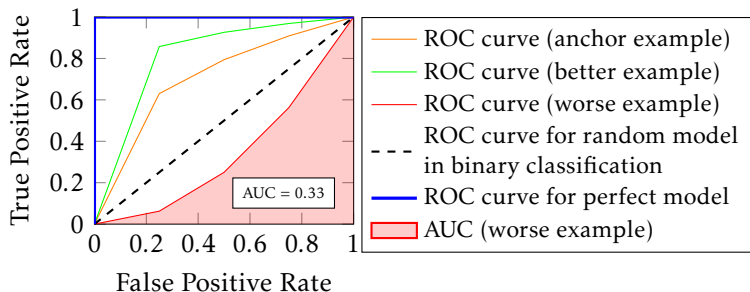- A model who is 100% correct has an AUC of 1.

AUC

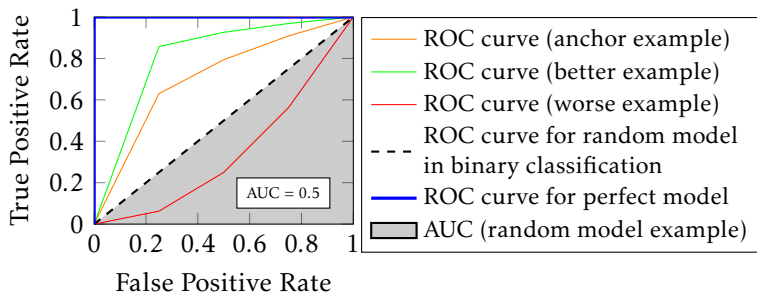**What is the AUC ?**

Figure: AUC for the worst ROC curve

Introduction
oo

Definitions
oooooooooooo●ooooooo

Contributions
ooooooooo

Applications
ooooooo

Conclusion
ooo

AUC

What is the AUC ?

Figure: AUC for the random model ROC curve

Introduction
○○

Definitions
○○○○○○○○○○○○○●○○○○○○

Contributions
○○○○○○○○

Applications
○○○○○○○

Conclusion
○○○

AUC
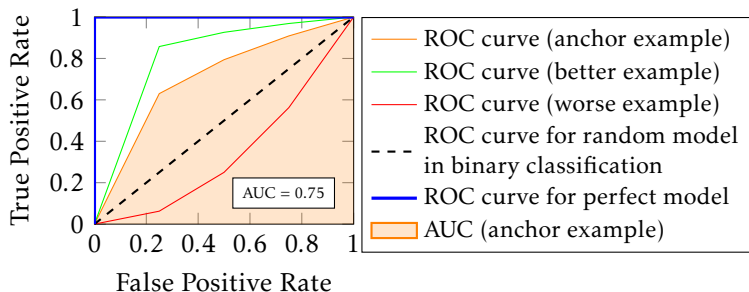
What is the AUC ?

Figure: AUC for the anchor ROC curve

Introduction
oo

Definitions
ooooooooooooooo●oooo

Contributions
ooooooooo

Applications
ooooooo

Conclusion
ooo

AUC
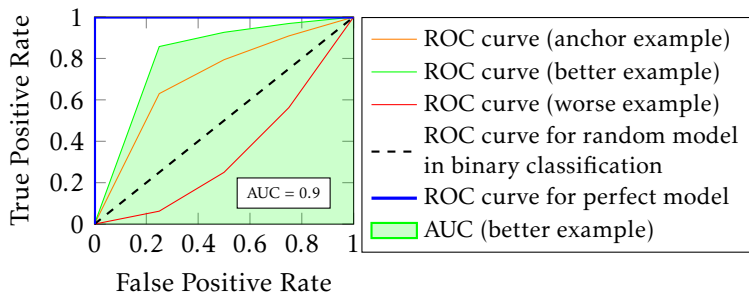
What is the AUC ?

Figure: AUC for the better ROC curve

Introduction
oo

Definitions
oooooooo●ooo

Contributions
ooooooooo

Applications
ooooooo

Conclusion
ooo

AUC

What is the AUC ?
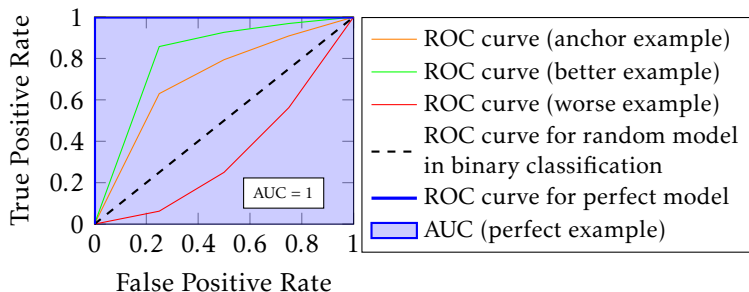
Figure: AUC for the perfect ROC curve

Outline

ROC and bipartite ranking

What is the link between ROC curves and bipartite ranking ?

- Different tasks require different metrics.
- **Classification** : accuracy, precision, recall, f1 score, etc.
- **Regression** : mean squared error, mean absolute error, etc.
- **None of these metrics take the rank into account.** They freeze the number of true/false positives/negatives for a **particular threshold** (usually 0.5).

ROC and bipartite ranking

The ROC curve **intrinsically embeds the information of the rank** by giving information on the confusion matrix for all possible thresholds.

Therefore, the analysis of the ROC curve is a **common solution** to assess the performance of a **ranking model**.

Outline

Contributions

The paper addresses the problem of **fairness** in **bipartite
ranking models**, which have different requirements than
classification models.

The authors came up with **two contributions** to **improve
fairness** of bipartite ranking models :

- AUC-based constraints
- ROC-based constraints

They show the **limitations** of the AUC-based constraints, and
how the ROC-based constraints **address** them.

Motivation

Fairness in ranking is important because it can have a
significant impact on the decision-making process. For
example, in the context of hiring, a biased ranking model can
lead to unfair hiring practices.

Outline

AUC-based constraints

Introduction
oo

Definitions
ooooooooooooooooooo

Contributions
ooooo●oo

Applications
ooooooo

Conclusion
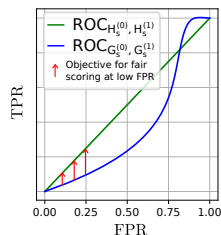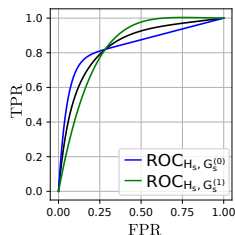ooo

Limits of AUC-based constraints
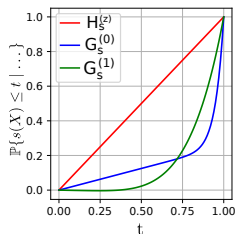
Figure: Illustrating the limitations of AUC-based fairness.

Outline

1 Introduction

2 Definitions

3 Contributions
   AUC-based constraints
   ROC based constraints

4 Applications

5 Conclusion

ROC based constraints

small change another change final change hopefully ok now it works

Outline

1. Introduction

2. Definitions

3. Contributions

4. Applications

5. Conclusion

Outline

Compas dataset

Outline

Introduction
oo

Definitions
ooooooooooooooooooo

Contributions
oooooooooo

Applications
ooooo●oo

Conclusion
ooo

Adult dataset

Outline

1. Introduction

2. Definitions

3. Contributions

4. Applications
   Compas dataset
   Adult dataset
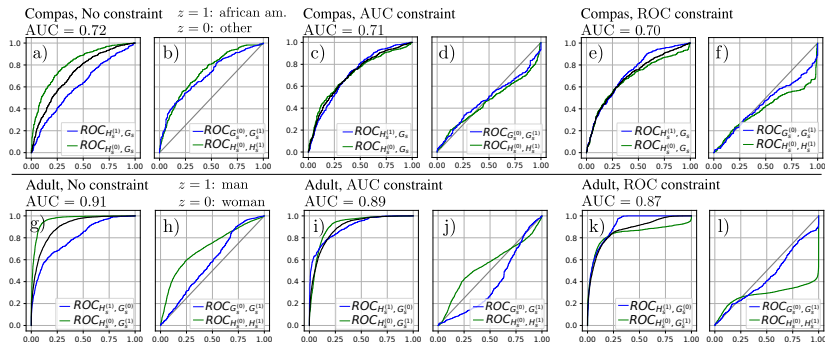   **Results**

5. Conclusion

Results

Figure: *ROC* curves on the test set of Adult and Compas for a score learned without and with fairness constraints. Black curves represent ROC_{H_s,G_s}. We also report the corresponding ranking performance AUC_{H_s,G_s}.

Outline

1. Introduction

2. Definitions

3. Contributions

4. Applications

5. Conclusion

Introduction
00

Definitions
00000000000000000000

Contributions
00000000

Applications
0000000

Conclusion
0●0

Conclusion

Bibliography

📄 Narasimhan, Harikrishna and Shivani Agarwal. "On the relationship between binary classification, bipartite ranking, and binary class probability estimation". In: *Advances in neural information processing systems* 26 (2013).

.