Fairness for AUC via Feature Augmentation

Hortense Fong
Yale University
USA

Vineet Kumar
Yale University
USA

Anay Mehrotra
Yale University
USA

Nisheeth K. Vishnoi
Yale University
USA

ABSTRACT

We study fairness in the context of classification where the performance is measured by the area under the curve (AUC) of the receiver operating characteristic. AUC is commonly used when both Type I (false positive) and Type II (false negative) errors are important. However, the same classifier can have significantly varying AUCs for different protected groups and, in real-world applications, it is often desirable to reduce such cross-group differences. We address the problem of how to select additional features to most greatly improve AUC for the disadvantaged group. Our results establish that the unconditional variance of features does not inform us about AUC fairness but class-conditional variance does. Using this connection, we develop a novel approach, fairAUC, based on feature augmentation (adding features) to mitigate bias between identifiable groups. We evaluate fairAUC on synthetic and real-world (COMPAS) datasets and find that it significantly improves AUC for the disadvantaged group relative to benchmarks maximizing overall AUC and minimizing bias between groups.

CCS CONCEPTS

• **Computing methodologies → Supervised learning by classification**.

KEYWORDS

Classification, Area under the ROC curve (AUC), feature augmentation, data collection and curation

ACKNOWLEDGMENTS