Title: Equalized odds

URL: https://en.wikipedia.org/wiki/Equalized_odds

PageID: 66205491

Categories: Category:Bias, Category:Computing and society, Category:Discrimination, Category:Ethics stubs, Category:Information ethics, Category:Machine learning, Category:Machine learning stubs, Category:Philosophy of artificial intelligence

---

Equalized odds , also referred to as conditional procedure accuracy equality and disparate mistreatment , is a measure of fairness in machine learning . A classifier satisfies this definition if the subjects in the protected and unprotected groups have equal true positive rate and equal false positive rate, satisfying the formula:

P (R = + | Y = y , A = a) = P (R = + | Y = y , A = b) y ∈ { + , – } ∀ a , b ∈ A {\displaystyle P(R=+|Y=y,A=a)=P(R=+|Y=y,A=b)\quad y\in \{+,-\}\quad \forall a,b\in A}

For example, A {\displaystyle A} could be gender, race, or any other characteristics that we want to be free of bias, while Y {\displaystyle Y} would be whether the person is qualified for the degree, and the output R {\displaystyle R} would be the school's decision whether to offer the person to study for the degree. In this context, higher university enrollment rates of African Americans compared to whites with similar test scores might be necessary to fulfill the condition of equalized odds, if the "base rate" of Y {\displaystyle Y} differs between the groups.

The concept was originally defined for binary-valued Y {\displaystyle Y} . In 2017, Woodworth et al. generalized the concept further for multiple classes.

See also

Fairness (machine learning)

Color blindness (racial classification)

References

This machine learning -related article is a stub . You can help Wikipedia by expanding it .

v

t

e

This article about ethics is a stub . You can help Wikipedia by expanding it .

v

t

e