HW 5

Neel Singh

11/8/2024

This homework is meant to give you practice in creating and defending a position with both statistical and philosophical evidence. We have now extensively talked about the COMPAS [1] data set, the flaws in applying it but also its potential upside if its shortcomings can be overlooked. We have also spent time in class verbally assessing positions both for an against applying this data set in real life. In no more than two pages [2] take the persona of a statistical consultant advising a judge as to whether they should include the results of the COMPAS algorithm in their decision making process for granting parole. First clearly articulate your position (whether the algorithm should be used or not) and then defend said position using both statistical and philosophical evidence. Your paper will be grade both on the merits of its persuasive appeal but also the applicability of the statistical and philosohpical evidence cited.

Student Response:

COMPAS is a predictive algorithm designed to assess the likelihood of recidivism for potential parolees. This algorithm analyzes many factors such as criminal history, age, and socioeconomic indicators to generate a risk score. COMPAS is designed to assist judges like yourself by providing an objective risk assessment to supplement your decision-making in parole hearings. This algorithm aims to enhance the fairness of parole decisions by reducing reliance on potentially subjective biases that may influence human judgment and by providing a consistent framework to evaluate individuals based on data-driven criteria. As this court's statistical consultant, I would highly recommend the use of COMPAS in all parole hearings as a supplementary tool.

However, I think it is important to inform you of a potential concern with COMPAS. Although it aims to offer an objective assessment, research has shown that COMPAS may exhibit biases in its predictions, particularly along racial lines. For instance, studies have indicated that it can have higher false-positive rates for certain minority groups. The false-positive rate is the frequency with which COMPAS mistakenly labels an individual as high risk when they are not. This means that individuals from these groups may be mistakenly assessed as higher risk than they actually are.

To directly address these concerns, the ethics of COMPAS can be examined through two principles: sufficiency and merit-based fairness.

Sufficiency is a statistical concept that COMPAS satisfies, helping ensure fairness across different racial groups. Mathematically, sufficiency means that the actual likelihood of recidivism, denoted Y, is conditionally independent of race S when we account for the COMPAS score $\hat{Y}$. In simpler terms, this means that once we know an individual's COMPAS score, their race does not influence their likelihood of reoffending.

To put it plainly, if two individuals have the same COMPAS score, they have the same risk of reoffending, regardless of race. This property is critical because it ensures that the score reflects a person's risk in a way that is consistent across demographic groups. By meeting the sufficiency criterion, COMPAS can provide a score that is equally predictive across racial groups.

While differing false-positive rates across racial groups have raised concerns about fairness in the use of COMPAS, it is important to note that these differences do not necessarily reflect unfairness in the algorithm's

[1] https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis
[2] knit to a pdf to ensure page count

predictions. Such variations may be due to differences in base recidivism rates—the underlying likelihood of reoffending within each group—rather than any inherent bias in the algorithm itself. In other words, the higher underlying risk of reoffending in one group may lead to a greater number of individuals within that group being classified as high risk. Some of these individuals will not reoffend, thereby increasing the false-positive rate for that group. This does not imply that the algorithm is biased against any particular group, but rather COMPAS reflects the influence of real, statistical differences in reoffending behavior across populations. Because COMPAS meets the principle of sufficiency, we can be confident that each risk score level $\hat{Y}$ consistently predicts recidivism risk Y across all groups. Thus, even if false-positive rates vary, the sufficiency principle assures that COMPAS's core predictions remain fair and reliable for all individuals.

In addition to sufficiency, COMPAS aligns with a merit-based fairness approach, which evaluates individuals based on relevant characteristics directly related to recidivism risk. Merit-based fairness holds that individuals should be judged on factors that genuinely indicate their own likelihood of reoffending, such as prior criminal history and behavioral indicators, rather than protected demographic characteristics like race. COMPAS does not account for an individual's race. It only sees information that is directly linked to recidivism risk. While some of these factors may correlate with race, this correlation may simply reflect a real difference in recidivism between racial groups. COMPAS assesses individuals based on their own risk-related attributes, providing a fair evaluation grounded in individual behavior rather than group characteristics.

This merit-based focus is crucial for supporting sufficiency, as it ensures that each high-risk score truly corresponds to a higher risk of recidivism. Because COMPAS evaluates individuals based on relevant risk factors, when an individual receives a high COMPAS score, that score reliably reflects a meaningful, data-supported increase in their risk of reoffending. This direct link between risk score and actual recidivism likelihood confirms that COMPAS's predictions are grounded in merit as individuals are rated high-risk because their background and behavior statistically indicate a greater chance of recidivism, rather than due to race-based characteristics.

In conclusion, COMPAS provides a consistent, objective framework for evaluating parole candidates based on data-driven criteria. By satisfying sufficiency, it ensures that each risk score is equally predictive across racial groups, addressing concerns about statistical fairness and supporting an unbiased assessment. Additionally, by aligning with merit-based fairness, COMPAS evaluates individuals on factors directly related to their risk of reoffending, grounding its predictions in the attributes that matter most for parole decisions.

These combined principles support COMPAS as a fair and effective tool for decision-making. COMPAS's data-driven approach reduces the influence of subjective biases, offering a standardized, equitable measure of risk that enhances the overall fairness and consistency of parole hearings. For these reasons, I believe COMPAS should be used as a supplementary tool, providing judges with a reliable, objective measure that supports informed and impartial decision-making in parole hearings.