# HW 5

Jiangyuan Yuan

11/08/2024

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*

I strongly advise against the integration of the COMPAS algorithm into the parole decision-making process. While COMPAS allegedly enhances efficiency and objectivity in predicting recidivism, substantial statistical deficiencies and profound philosophical concerns render its application both morally and practically untenable. This position is supported by an in-depth analysis of COMPAS's performance against established fairness metrics and its alignment with core ethical principles of justice and individual rights.

COMPAS demonstrates significant disparities in its classification rates across protected and unprotected classes. The Disparate Impact metric assesses whether the rate of positive classifications (i.e., high-risk assessments) is proportionately similar between groups. For COMPAS, the value is roughly equal to 0.5916, falling well below the threshold of 0.8, which indicates that individuals from the protected class are disproportionately labeled as high-risk compared to their counterparts. This violation suggests that COMPAS may be perpetuating systemic biases rather than mitigating them. Similarly, Statistical Parity measures the absolute difference in positive classification rates between groups. For COMPAS, the difference of 0.2402 is substantial, further highlighting its unfair treatment of protected classes. Such significant disparities undermine the algorithm's claim to impartiality and fairness, raising red flags about its equitable application in parole decisions. Moreover, COMPAS fails to satisfy the Equalized Odds criterion, which requires that both false positive rates and true positive rates are equal across groups. The observed discrepancies of $|0.218|$ in FPR and $|0.1974|$ in TPR indicate that COMPAS disproportionately misclassifies individuals from the protected class. Specifically, a higher FPR means that more innocent individuals from the protected class are incorrectly labeled as high-risk, leading to an unjust denial of parole. Conversely, unequal TPRs imply that high-risk individuals are not uniformly identified, potentially compromising public safety if such an algorithm allowed certain individuals to be granted parole.

The Impossibility Theorem asserts that no single classifier can simultaneously satisfy all fairness criteria. COMPAS's inability to reconcile these metrics exemplifies the inherent trade-offs in algorithmic fairness. While striving for one fairness measure, COMPAS inadvertently compromises others, revealing its fundamental limitations in achieving comprehensive fairness. Additionally, the trade-off between accuracy and fairness becomes evident, as imposing strict fairness constraints typically results in reduced predictive accuracy. However, COMPAS's bias-induced inaccuracies far outweigh any marginal gains in efficiency, making it an unreliable tool for parole decisions.

[1] https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis

[2] knit to a pdf to ensure page count

From a philosophical standpoint, John Rawls's Difference Principle advocates for structuring societal institutions to benefit the least advantaged members. By disproportionately classifying individuals from protected classes as high-risk, COMPAS contradicts this principle. Instead of leveling the playing field, COMPAS exacerbates existing inequalities, perpetuating the systemic disadvantages faced by marginalized groups. This misalignment with Rawlsian justice underscores the ethical unfitness of COMPAS for parole decisions, as it fails to prioritize the well-being of the most vulnerable members of society.

Another philosophical criticism of using COMPAS in the parole process is Immanuel Kant's ethical framework, which emphasizes treating individuals as ends in themselves. COMPAS's reliance on predictive analytics reduces defendants to data points, stripping them of their intrinsic dignity and individuality. The inability to explain or understand COMPAS's reasoning process further compounds this ethical violation, as it prevents meaningful engagement with each defendant's unique circumstances. By operating as a black-box, COMPAS undermines the Kantian imperative of respecting each person's inherent worth and autonomy, making its use in parole decisions morally indefensible.

This lack of transparency from acting as a black box hinders the ability to audit and understand the factors influencing parole decisions, undermining accountability. Judges must adequately justify their decisions or identify potential biases within the algorithm with clear insights into how specific inputs influence outcomes. This opacity diminishes trust in the judicial system and impedes efforts to rectify any underlying biases that COMPAS may perpetuate.

The ethical deployment of algorithms in sensitive areas like criminal justice hinges on public trust. COMPAS's biased outcomes and lack of transparency erode confidence in the judicial process, fostering skepticism and undermining the legitimacy of parole decisions. Public trust is foundational to the effectiveness of any justice system, and the use of COMPAS threatens to ruin this trust by perpetuating perceived injustices. Additionally, the opaque nature of COMPAS shifts accountability from human judges to the algorithm itself, diluting responsibility for unjust outcomes. When an incomprehensible model influences parole decisions, holding any party accountable for wrongful denials or approvals becomes challenging. This diffusion of responsibility undermines the ethical duty of judges to make informed and fair decisions, further justifying the exclusion of COMPAS from parole considerations.

In conclusion, the statistical deficiencies and philosophical misalignments of the COMPAS algorithm present compelling reasons to exclude it from parole decision-making processes. Its failure to satisfy essential fairness metrics and ethical violations related to transparency, accountability, and respect for individuals render COMPAS an unsuitable and morally problematic tool for such critical applications. Parole decisions should remain grounded in transparent, fair, and ethically accountable human judgment but allowed to be supportyed by tools that demonstrably uphold statistical fairness and ethical integrity. Upholding these principles is essential to maintaining the justice system's legitimacy, ensuring that parole decisions are both just and equitable.