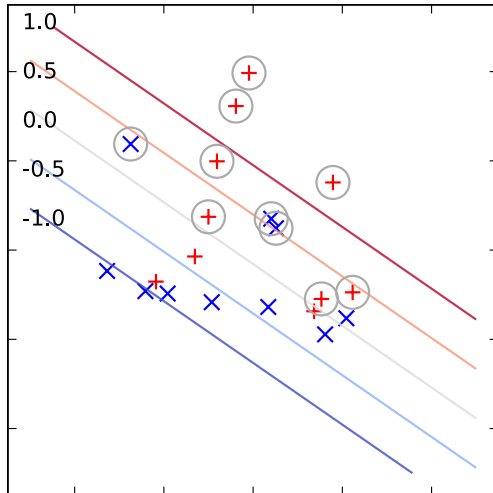


TAKE HOME, OPEN BOOK BONUS QUESTION:
 FAIRNESS AND CAUSALITY IN ML
 (20PT EXTRA CREDIT, APPLIED POST CURVE)

Don't feel confident in your answers? Feel confident but have nothing planned for the weekend? Take these two pages home and return your *individual* solution in class on Tuesday 10/29/19 for a chance at 20 extra points (only to be applied after any and all grade curving).



Displayed on the left are past applicants for jobs at a federal department. Suppose that $+$ are *qualified* job applicants and $\times$ are *unqualified* applicants from historical data. The features are performance in two competency tests. Circled points correspond to *androids* and non-circled points are *humans*. The true population is 50% human and 50% android.

There will be 10 new applicants applying for federal jobs in our department tomorrow. We want to build a classifier that will tell us who to hire. The new applicants will be drawn from the same population that past applicants were drawn from.

- (a) (3 pts.) We want to hire approximately half of these applicants; *i.e.* our goal is $P(\text{hire}) \approx 0.5$. Choose a single threshold that is closest to this target. _____
- (a) How many qualified applicants do we expect to hire? _____
- (b) What percent of qualified androids do we expect to hire? _____
- (c) What percent of qualified humans do we expect to hire? _____
- (b) (2 pts.) The Human Nondiscrimination Act of 2036 maintains that all federal employers must provide equal opportunities to androids and humans. What two thresholds should you choose that work for our targets so that the fraction of qualified humans that we hire is the same as for androids (on average)?
 Human: _____ Android: _____ (Please try to get as close to the target as you can.)
- How many qualified applicants do we expect to hire? _____
- (c) (2 pts.) The Human Non-Irrelevancy Act of 2045 amends the law to say that all federal employers must (on average) employ the same number of androids and humans, to reflect that 50-50 split in the US combined population. What two thresholds should you choose that work for our targets and hire on average the same number of humans and androids? _____ and _____

How many qualified applicants do we expect to hire? _____

- (d) **(5 pts.)** The android C-3PO didn't get a job even though it tested better than a human who ended up being hired instead. C-3PO sued the government and won! Federal employers are no longer allowed to have two different thresholds for hiring. At the same time, the equal-proportion quotas of the Human Non-Irrelevancy Act of 2045 remain the law of the land. Damned if we do and damned if we don't?! How can we make sure not to get in trouble and still hire $\approx 50\%$ of applicants? Note that: **A.** Our classifier has to hire $\approx 50\%$ of applicants; **B.** $\approx 50\%$ of them must be human; **C.** We cannot use humanness as input to decide upon. Come up with a solution and report how many qualified applicants you expect to hire.

- (e) **(8 pts.)** Applicants are deemed "qualified" if, were they hired, they would end up successfully completing 95% of tasks given to them within the first 6 months. We are not getting enough qualified candidates to keep our department running! What can we do?

Nathan and Xiaojie offer their expert consulting help (for a significant fee). They build the above classification model to predict qualification based on test scores and come up with a truly inspired solution: to get more qualified candidates, we should provide free prep classes for our competency tests to every applicant. The prep classes are *sure* to increase any applicant's score. Then, more of our applicants will have higher test scores and therefore more of our applicants will be qualified, as predicted by the model, which was based on *very solid* ML.

Will Nathan and Xiaojie's plan work? Why might it or why might it not? Try to come up with made up scenarios for both cases.