

Go here:

<http://research.google.com/bigpicture/attacking-discrimination-in-ml/>

### “Simulating Loan Thresholds”

1. What threshold maximizes the number of correct decisions? Does that same threshold also minimize the number of incorrect decisions?
2. What threshold maximizes profit? (See the paragraph beginning “Another goal” for details.) You should see a profit of at least 16600.
3. Are your answers to #1 and #2 the same? Why or why not? Use the language of “False Positive” and “False Negative”.

### Classification and Discrimination

**Don’t click the red buttons** until directed. Instead, just drag the threshold sliders.

4. Describe, in words, how “blue people” and “orange people” differ. Specifically, the text says: *“In this case, the distributions of the two groups are slightly different, even though blue and orange people are equally likely to pay off a loan.”* In what way are the distributions different?
5. Adjust the thresholds to **maximize the profit** that the bank gets for each group individually. (a) What are the resulting thresholds for the two groups? (b) At those thresholds, which group is *offered* more loans? (c) Of the loans that were offered, which group *repays* more of those loans? (d) What is the resulting profit for the bank? (e) Is that fair? (explain.) (Check your answer by clicking on the red “Max Profit” button.)
6. Adjust the thresholds to maximize profit **under the constraint that blue and orange must have the same threshold**. (This is tricky in the UI; hint: it must be between the two thresholds you found in the previous exercise.) Answer the same 5 questions (a-e) as in the previous exercise. Then check your answer by clicking the “Group Unaware” button.
7. Starting at the “Group Unaware” thresholds, **change the threshold on the orange group** to increase its Positive Rate to **match the Positive Rate of the blue group**. (a) What happens to the True Positive Rate? (b) What happens to the bank’s profit? (c) Is that good? Explain.
8. Explore the “Demographic Parity” and “Equal Opportunity” objectives by clicking their buttons. The authors of the interactive visualization seem to have a favorite among these objectives... what reasons might someone have to agree or disagree with their assessment?