Open in Colab

Exploring the ProPublica COMPAS Analysis

In this notebook you'll get a chance to examine the data used in the ProPublica story yourself.

Disclaimer: Please don't over interpret what you find in the data. We know from our discussions that methodology is key to being able to properly interpret findings. Our goal here will be to reproduce results from the readings we did for last class.

First, we'll download and parse the data into a data frame.

In [2]:

```
import pandas as pd

!wget https://raw.githubusercontent.com/propublica/compas-analysis/master/compas-scores-two-years.csv
df = pd.read_csv('compas-scores-two-years.csv')
df
```

```
--2019-10-24 03:07:49--  https://raw.githubusercontent.com/propublica/compas-analysis/master/compas-scores-two-years.csv
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 151.101.0.133, 151.101.64.133, 151.101.128.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|151.101.0.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2546489 (2.4M) [text/plain]
Saving to: ‘compas-scores-two-years.csv’

compas-scores-two-y 100%[===================>]   2.43M  --.-KB/s    in 0.09s

2019-10-24 03:07:49 (26.4 MB/s) - ‘compas-scores-two-years.csv’ saved [2546489/2546489]
```

Out[2]:

	id	name	first	last	compas_screening_date	sex	dob	age	age_cat	race	juv_fel_count	dec
0	1	miguel hernandez	miguel	hernandez	2013-08-14	Male	1947-04-18	69	Greater than 45	Other	0	
1	3	kevon dixon	kevon	dixon	2013-01-27	Male	1982-01-22	34	25 - 45	African-American	0	
2	4	ed philo	ed	philo	2013-04-14	Male	1991-05-14	24	Less than 25	African-American	0	
3	5	marcu brown	marcu	brown	2013-01-13	Male	1993-01-21	23	Less than 25	African-American	0	
4	6	bouthy pierrelouis	bouthy	pierrelouis	2013-03-26	Male	1973-01-22	43	25 - 45	Other	0	
5	7	marsha miles	marsha	miles	2013-11-30	Male	1971-08-22	44	25 - 45	Other	0	
6	8	edward riddle	edward	riddle	2014-02-19	Male	1974-07-23	41	25 - 45	Caucasian	0	
7	9	steven stewart	steven	stewart	2013-08-30	Male	1973-02-25	43	25 - 45	Other	0	
8	10	elizabeth thieme	elizabeth	thieme	2014-03-16	Female	1976-06-03	39	25 - 45	Caucasian	0	
9	13	bo bradac	bo	bradac	2013-11-04	Male	1994-06-10	21	Less than 25	Caucasian	0	
		benjamin					1988-					

	id	name	first	last	compas_screening_date	sex	dob	age	age_cat	race	juv_fel_count	dec
10	14	benjamin franc	benjamin	franc	2013-11-26	Male	1988-06-09	27	25 - 45	Caucasian	0	
11	15	ellyaher lanza	ellyaher	lanza	2013-10-03	Male	1992-08-18	23	Less than 25	African-American	0	
12	16	kortney coleman	kortney	coleman	2013-01-01	Female	1978-08-22	37	25 - 45	Caucasian	0	
13	18	jarrod turbe	jarrod	turbe	2013-10-09	Male	1974-12-02	41	25 - 45	African-American	0	
14	19	craig gilbert	craig	gilbert	2013-10-30	Female	1968-06-14	47	Greater than 45	Caucasian	0	
15	20	samuel seraphin	samuel	seraphin	2014-06-03	Male	1985-03-25	31	25 - 45	African-American	0	
16	21	mario hernandez	mario	hernandez	2014-03-24	Male	1979-01-25	37	25 - 45	Hispanic	0	
17	22	darrious davis	darrious	davis	2013-12-22	Male	1990-06-22	25	25 - 45	African-American	0	
18	23	neil heckart	neil	heckart	2013-11-17	Male	1984-12-24	31	25 - 45	Caucasian	0	
19	24	michael lux	michael	lux	2014-11-15	Male	1985-01-08	31	25 - 45	Caucasian	0	
20	25	columbus wilson	columbus	wilson	2014-05-02	Male	1951-06-28	64	Greater than 45	African-American	0	
21	26	vandivuiet williams	vandivuiet	williams	2013-04-18	Male	1994-11-29	21	Less than 25	African-American	0	
22	27	nelson avalo	nelson	avalo	2014-10-16	Male	1988-08-06	27	25 - 45	Caucasian	0	
23	28	janel denicola	janel	denicola	2013-11-22	Female	1995-03-22	21	Less than 25	Caucasian	0	
24	30	dominic pabon	dominic	pabon	2013-02-08	Male	1992-01-23	24	Less than 25	Hispanic	0	
25	32	russell sottile	russell	sottile	2013-01-25	Male	1973-01-10	43	25 - 45	Caucasian	0	
26	33	andre ashley	andre	ashley	2013-05-11	Male	1983-08-24	32	25 - 45	Other	0	
27	37	deandrae counts	deandrae	counts	2013-05-06	Male	1989-02-08	27	25 - 45	African-American	0	
28	38	victoria soltau	victoria	soltau	2013-03-18	Female	1979-09-03	36	25 - 45	Caucasian	0	
29	39	najee sapp	najee	sapp	2013-02-20	Male	1989-04-27	26	25 - 45	African-American	0	
...	...	...	...	...	...	...	...	...	...	...	...	...
7184	10962	yolani moratz	yolani	moratz	2013-08-08	Female	1979-11-05	36	25 - 45	Caucasian	0	
7185	10963	paul wyatt	paul	wyatt	2013-11-06	Male	1973-09-19	42	25 - 45	Caucasian	0	
7186	10964	terrence brown	terrence	brown	2013-02-23	Male	1979-06-08	36	25 - 45	African-American	0	
7187	10965	anthony fields	anthony	fields	2013-03-13	Male	1959-02-20	57	Greater than 45	Caucasian	0	

	id	name	first	last	compas_screening_date	sex	dob	age	age_cat	race	juv_fel_count	dec
7188	10966	shameel koya	shameel	koya	2013-03-04	Male	1979-10-12	36	25 - 45	Caucasian	0	
7189	10967	george bedward	george	bedward	2014-05-31	Male	1981-05-26	34	25 - 45	African-American	0	
7190	10969	eric sparks	eric	sparks	2013-01-11	Male	1991-07-13	24	Less than 25	African-American	0	
7191	10971	eugene drogus	eugene	drogus	2014-01-07	Male	1948-10-17	67	Greater than 45	Caucasian	0	
7192	10972	matt munoz	matt	munoz	2013-09-12	Male	1984-03-22	32	25 - 45	Caucasian	0	
7193	10975	warren aiken	warren	aiken	2013-09-05	Male	1990-09-30	25	25 - 45	African-American	0	
7194	10976	arleen martin	arleen	martin	2014-12-19	Female	1985-08-14	30	25 - 45	Caucasian	0	
7195	10977	adrian williams	adrian	williams	2013-09-18	Male	1981-03-09	35	25 - 45	African-American	0	
7196	10979	angelita diaz	angelita	diaz	2013-09-06	Male	1972-07-19	43	25 - 45	African-American	0	
7197	10980	jarvis yates	jarvis	yates	2014-02-27	Male	1987-08-27	28	25 - 45	African-American	0	
7198	10981	orett harrison	orett	harrison	2013-12-25	Male	1984-03-31	32	25 - 45	African-American	0	
7199	10982	austin harris	austin	harris	2013-10-01	Male	1992-07-07	23	Less than 25	Caucasian	0	
7200	10984	shantrina stfort	shantrina	stfort	2013-11-05	Female	1995-06-06	20	Less than 25	African-American	0	
7201	10985	kyle miller	kyle	miller	2014-01-22	Male	1986-04-08	30	25 - 45	African-American	0	
7202	10987	ceasar gomez	ceasar	gomez	2013-03-31	Male	1990-02-07	26	25 - 45	Hispanic	0	
7203	10988	luis fernandez	luis	fernandez	2013-10-27	Male	1971-09-19	44	25 - 45	Hispanic	0	
7204	10989	rodney montgomery	rodney	montgomery	2013-12-28	Male	1985-09-28	30	25 - 45	African-American	0	
7205	10990	christopher tun	christopher	tun	2013-05-28	Male	1992-04-28	23	Less than 25	Caucasian	0	
7206	10992	alexander vega	alexander	vega	2013-05-10	Male	1994-07-15	21	Less than 25	Caucasian	0	
7207	10994	jarred payne	jarred	payne	2014-05-10	Male	1985-07-31	30	25 - 45	African-American	0	
7208	10995	raheem smith	raheem	smith	2013-10-20	Male	1995-06-28	20	Less than 25	African-American	0	
7209	10996	steven butler	steven	butler	2013-11-23	Male	1992-07-17	23	Less than 25	African-American	0	
7210	10997	malcolm simmons	malcolm	simmons	2014-02-01	Male	1993-03-25	23	Less than 25	African-American	0	
7211	10999	winston gregory	winston	gregory	2014-01-14	Male	1958-10-01	57	Greater than 45	Other	0	

	id	name	first	last	compas_screening_date	sex	dob	age	age_cat	race	juv_fel_count	dec
7212	11000	farrah jean	farrah	jean	2014-03-09	Female	1982-11-17	33	25 - 45	African-American	0	
7213	11001	florencia sanmartin	florencia	sanmartin	2014-06-30	Female	1992-12-18	23	Less than 25	Hispanic	0	

7214 rows × 53 columns

In the ProPublica dataset they only used data where the "days_b_screening_arrest" feature was in the range [-30, 30].

In [0]:

```
# filter these based on propublica analysis (not sure why this doesn't match
https://fairmlbook.org/classification.html)
df = df[(df['days_b_screening_arrest'] <= 30) | (df['days_b_screening_arrest'] >= -30)]
```

Reproducing Calculations of False Positive Rate and Positive Predictive Value

From a statistical point of view (but not necessarily a social justice point of view) the debate between Propublica and NorthPointe boiled down to what is the rate way to measure bias in an algorithm. As you saw in the first part of the assignmnet, ProPublica used the evidence that the false positive rate differed between blacks and whites as evidence of bias. Northpointe argued used the fact that the positive predictive values across the two groups were the same as evidence *against* bias.

In Northpointe's report, they have a table which lists various statistics for blacks versus whites using the COMPAS risk scores as the predictor. Here is the relevant information from their report.

	Race	$\hat{y} = 1$	True positive rate	False Positive Rate	Positive Predictive Value	Negative Predictive Value
white	$decile \geq 1$	1.00	1.00	0.39	1.00	
		$decile \geq 2$	0.85	0.64	0.46	0.79
		$decile \geq 3$	0.74	0.47	0.5	0.76
		$decile \geq 4$	0.64	0.35	0.54	0.74
		$decile \geq 5$	0.52	0.23	0.59	0.71
		$decile \geq 6$	0.41	0.15	0.64	0.69
		$decile \geq 7$	0.29	0.09	0.68	0.66
		$decile \geq 8$	0.20	0.05	0.71	0.65
		$decile \geq 9$	0.12	0.03	0.70	0.63
		$decile \geq 10$	0.05	0.01	0.70	0.61
black	$decile \geq 1$	1.00	1.00	0.51	1.00	
		$decile \geq 2$	0.95	0.83	0.55	0.77
		$decile \geq 3$	0.89	0.68	0.58	0.73
		$decile \geq 4$	0.81	0.56	0.60	0.69
		$decile \geq 5$	0.72	0.45	0.63	0.65
		$decile \geq 6$	0.63	0.34	0.66	0.62
		$decile \geq 7$	0.51	0.25	0.69	0.59
		$decile \geq 8$	0.39	0.16	0.72	0.57
		$decile \geq 9$	0.26	0.09	0.74	0.54
		$decile \geq 10$	0.12	0.03	0.79	0.51

Notebook Exercise 1

Write code to reproduce the table above. Here are some hints to help you.

- If you're fuzzy on what each of these statistic means (false positive rate, true positive rate, etc.), consider checking out Binary

Diagnostic Tests.

- You'll want to use the column `df['two_year_recid']` as your indicator of true positive versus true negative (positive means that the person recidivated).
- To select narrow the data frame to just contain people of a particular race you can use the following snippet (`race` would be a string that is either "Caucasian" or "African-American").

```
df_for_race = df[df['race'] == race]
```

- To generate a particular row of the table, you'll want to loop over all possible thresholds where the model would predict recidivate ($\hat{y} = 1$).
- You can count the number of elements in a Pandas series that satisfy some criterion using the following technique. For instance, if we wanted to calculate the number of elements in "some_column" that are greater than 0 and less than 30, we could use the following code.

```
((df['some column'] > 0).sum() & (df['some column'] < 30).sum()
```

- It's up to you how you want to generate the table. You can simply print out the values within a loop as you compute them, or you could populate a data frame with your calculations and then plot them (this is what we did in the solution).

In [23]:

```
# ***Solution***

# we're going to create a data frame to hold all of the results.  Think of this
# as a represntation of the
results = pd.DataFrame(columns=['race', 'decile >=', 'true_positive_rate', 'false_positive_rate', '
positive_predictive_value', 'negative_predictive_value'])

for race in ['Caucasian', 'African-American']:
    df_for_race = df[df['race'] == race]
    y = df_for_race['two_year_recid']
    for thresh in range(1, 11):
        yhat = df_for_race['decile_score'] >= thresh
        true_positive_rate = ((y == 1) & (yhat == 1)).sum() / (y == 1).sum()
        false_positive_rate = ((y == 0) & (yhat == 1)).sum() / (y == 0).sum()
        if (yhat == 1).sum() == 0:
            positive_predictive_value = float('nan')
        else:
            positive_predictive_value = ((y == 1) & (yhat == 1)).sum() / (yhat == 1).sum()

        if (yhat == 0).sum() == 1:
            negative_predictive_value = float('nan')
        else:
            negative_predictive_value = ((y == 0) & (yhat == 0)).sum() / (yhat == 0).sum()

        results = results.append({'race': race,
                                  'decile >=': str(thresh),
                                  'true_positive_rate': true_positive_rate,
                                  'false_positive_rate': false_positive_rate,
                                  'positive_predictive_value': positive_predictive_value,
                                  'negative_predictive_value': negative_predictive_value}, ignore_i
dex=True)
results
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:18: RuntimeWarning: invalid value
encountered in long_scalars
```

Out[23]:

	race	decile >=	true_positive_rate	false_positive_rate	positive_predictive_value	negative_predictive_value
0	Caucasian	1	1.000000	1.000000	0.402439	NaN
1	Caucasian	2	0.852665	0.638987	0.473318	0.784404
2	Caucasian	3	0.736677	0.469388	0.513848	0.749503
3	Caucasian	4	0.642633	0.348346	0.554054	0.730284
4	Caucasian	5	0.526646	0.231527	0.605042	0.706796
5	Caucasian	6	0.410658	0.144265	0.657191	0.683146
6	Caucasian	7	0.294671	0.090077	0.687805	0.657012
7	Caucasian	8	0.203762	0.052780	0.722222	0.638520

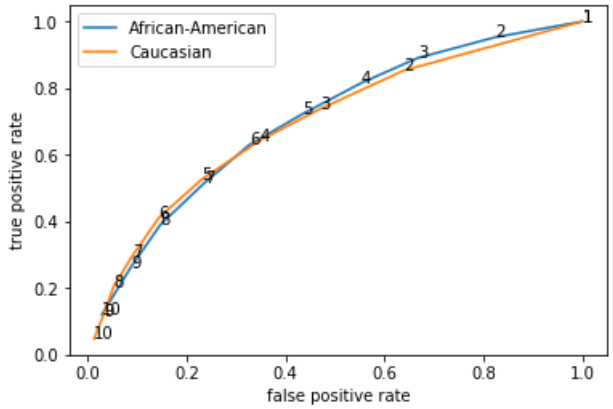
	race	decile >=	true_positive_rate	false_positive_rate	positive_predictive_value	negative_predictive_value
8	Caucasian	9	0.118077	0.032372	0.710692	0.619648
9	Caucasian	10	0.047022	0.013371	0.703125	0.605877
10	African-American	1	1.000000	1.000000	0.528414	NaN
11	African-American	2	0.952381	0.823141	0.564542	0.768229
12	African-American	3	0.890316	0.667266	0.599208	0.730263
13	African-American	4	0.815409	0.553357	0.622803	0.683486
14	African-American	5	0.721776	0.434053	0.650748	0.644809
15	African-American	6	0.629749	0.327338	0.683111	0.618523
16	African-American	7	0.515784	0.236811	0.709345	0.584481
17	African-American	8	0.390048	0.147482	0.747692	0.555035
18	African-American	9	0.260567	0.086930	0.770570	0.524269
19	African-American	10	0.119315	0.028777	0.822878	0.496020

In [25]:

```
import matplotlib.pyplot as plt

fig, ax = plt.subplots()
legend_strings = []
for race, df_by_race in results.groupby('race'):
    plt.plot(df_by_race['false_positive_rate'], df_by_race['true_positive_rate'])
    for _, row in df_by_race.iterrows():
        ax.annotate(str(row[1]), (row[3], row[2]))
    legend_strings.append(race)

plt.legend(legend_strings)
plt.xlabel('false positive rate')
plt.ylabel('true positive rate')
plt.show()
```

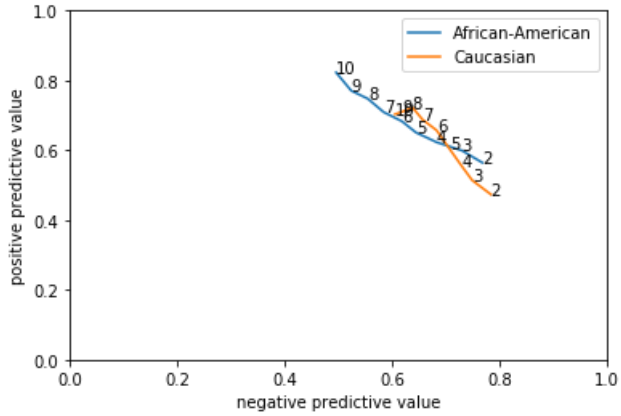


Notice how for a given threshold (annotated as text), the curves have vastly different false positive rates.

In [26]:

```
fig, ax = plt.subplots()
legend_strings = []
for race, df_by_race in results.groupby('race'):
    plt.plot(df_by_race['negative_predictive_value'], df_by_race['positive_predictive_value'])
    for _, row in df_by_race.iterrows():
        ax.annotate(str(row[1]), (row[5], row[4]))
    legend_strings.append(race)

plt.legend(legend_strings)
plt.xlim([0, 1])
plt.ylim([0, 1])
plt.xlabel('negative predictive value')
plt.ylabel('positive predictive value')
plt.show()
```



Notice how for a threshold of 4 or 5, the positive predictive values are almost identical.

Sanity Check Using Sklearn

We can compare our calculations to the ROC curve (false positive rate versus true postive rate).

In [7]:

```
from sklearn import metrics

for race in ['African-American', 'Caucasian']:
    df_filtered = df[df['race'] == race]
    y = df_filtered['two_year_recid']
    scores = df_filtered['decile_score']
    fpr, tpr, thresholds = metrics.roc_curve(y, scores)
    plt.plot(fpr, tpr)
plt.legend(['Caucasian', 'African-American'])
plt.xlabel('false positive rate')
plt.ylabel('true positive rate')
plt.show()
```