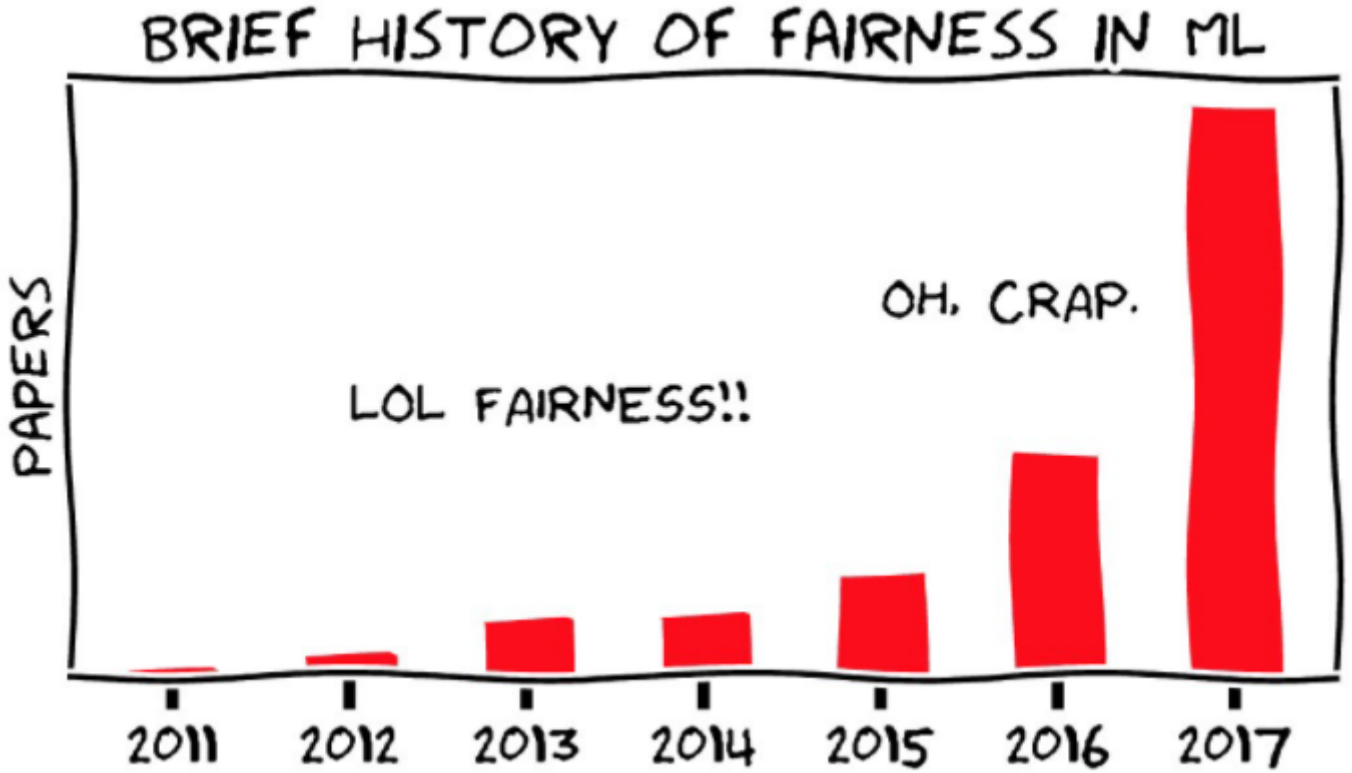
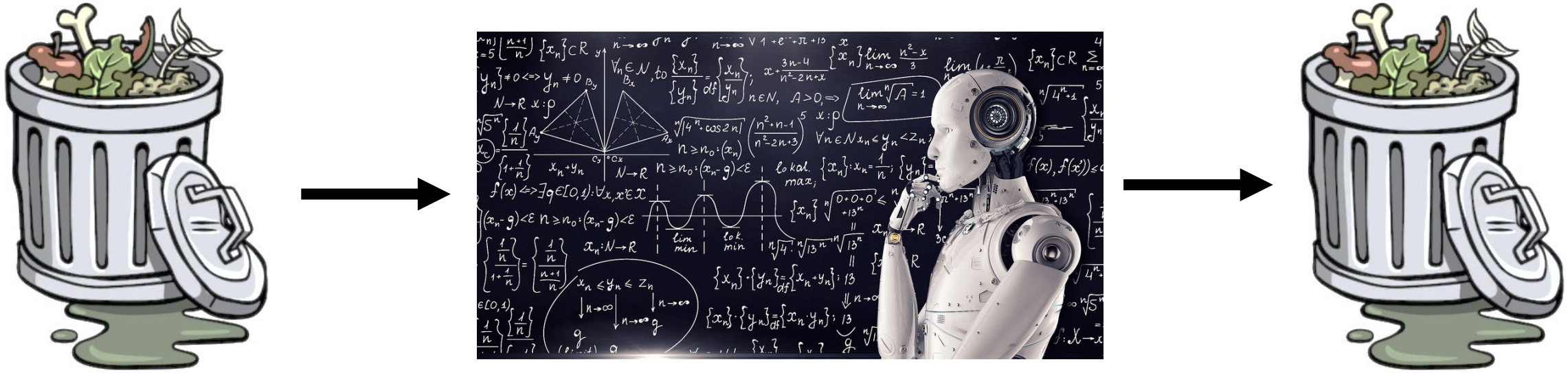
# Predictive Analytics Pt. II




Lab 7

# Predictive Analytics Pt. II



Garbage in, garbage out

Lab 7

# Agenda

- Revisiting the COMPAS data
- Bias in predictive models – sources and strategies
- Gentle introduction to neural networks

- Implementing a neural network in R

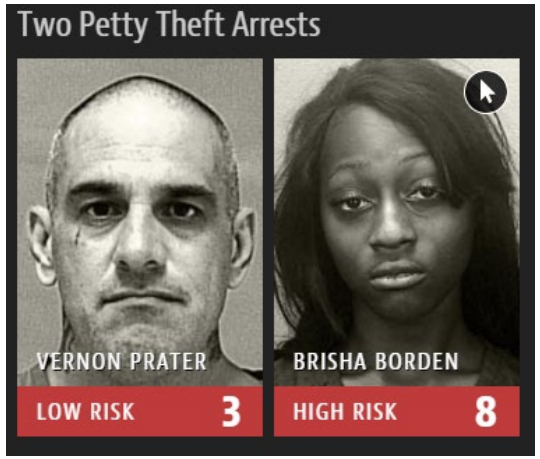
[Video]

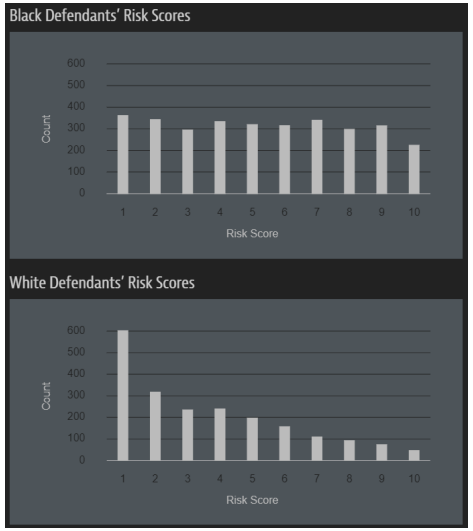
# Machine Bias

## There's software used across the country to predict future criminals. And it's biased against blacks.

*by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica*

May 23, 2016




### Prediction Fails Differently for Black Defendants

| | WHITE | AFRICAN AMERICAN |
|---|---|---|
| Labeled Higher Risk, But Didn't Re-Offend | 23.5% | 44.9% |
| Labeled Lower Risk, Yet Did Re-Offend | 47.7% | 28.0% |

***Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes.*** *(Source: ProPublica analysis of data from Broward County, Fla.)*

In [3]:
```
%%R
df <- dplyr::select(raw_data, age, c_charge_degree, race, age_cat, score_text, sex, priors_count,
                    days_b_screening_arrest, decile_score, is_recid, two_year_recid, c_jail_in, c_jail_out) %>%
        filter(days_b_screening_arrest <= 30) %>%
        filter(days_b_screening_arrest >= -30) %>%
        filter(is_recid != -1) %>%
        filter(c_charge_degree != "O") %>%
        filter(score_text != 'N/A')
nrow(df)
```

```
[1] 6172
```

Higher COMPAS scores are slightly correlated with a longer length of stay.

In [4]:
```
%%R
df$length_of_stay <- as.numeric(as.Date(df$c_jail_out) - as.Date(df$c_jail_in))
cor(df$length_of_stay, df$decile_score)
```

```
[1] 0.2073297
```

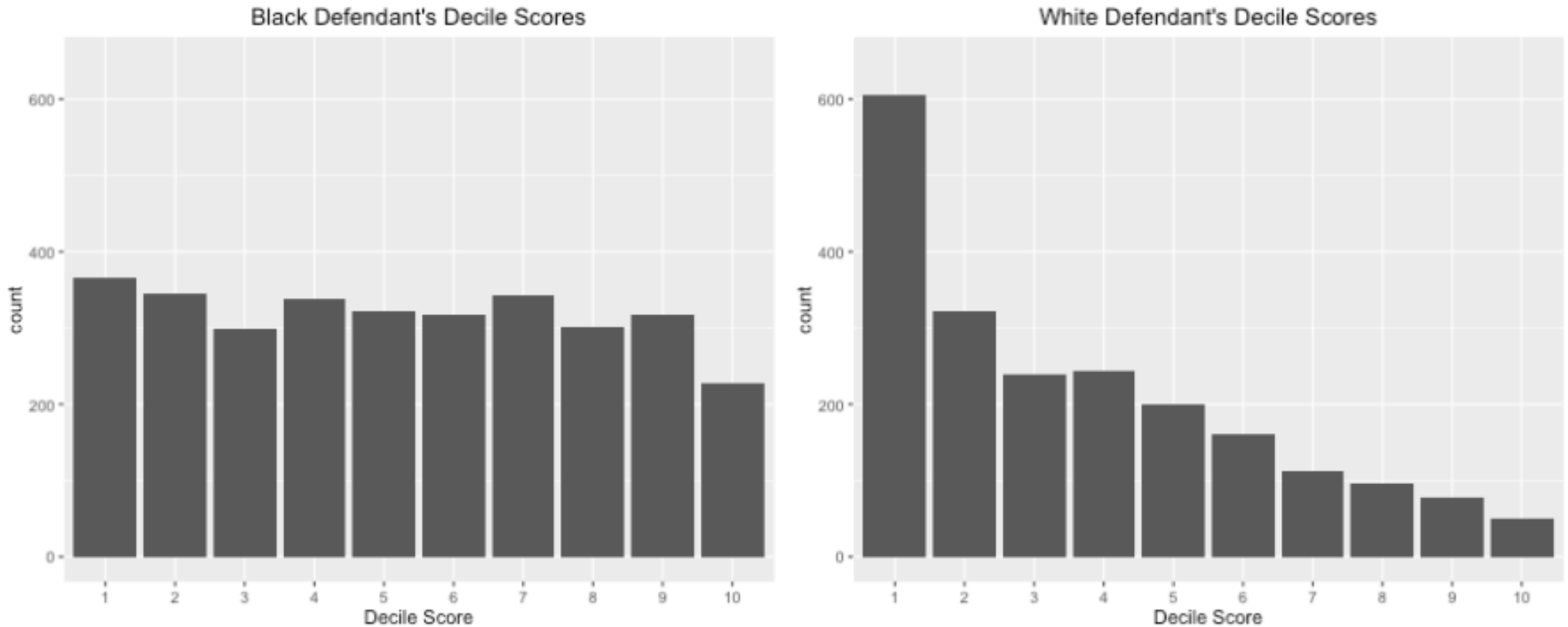
After filtering we have the following demographic breakdown:

In [5]:
```
%%R
summary(df$age_cat)
```

```
      25 - 45 Greater than 45    Less than 25
         3532            1293            1347
```

In [14]:

```
%%R -w 900 -h 363 -u px
library(grid)
library(gridExtra)
pblack <- ggplot(data=filter(df, race =="African-American"), aes(ordered(decile_score))) +
          geom_bar() + xlab("Decile Score") +
          ylim(0, 650) + ggtitle("Black Defendant's Decile Scores")
pwhite <- ggplot(data=filter(df, race =="Caucasian"), aes(ordered(decile_score))) +
          geom_bar() + xlab("Decile Score") +
          ylim(0, 650) + ggtitle("White Defendant's Decile Scores")
grid.arrange(pblack, pwhite,  ncol = 2)
```



# Racial Bias in Compas

After filtering out bad rows, our first question is whether there is a significant difference in Compas scores between races. To do so we need to change some variables into factors, and run a logistic regression, comparing low scores to high scores.

In [16]:

```
%%R
df <- mutate(df, crime_factor = factor(c_charge_degree)) %>%
      mutate(age_factor = as.factor(age_cat)) %>%
      within(age_factor <- relevel(age_factor, ref = 1)) %>%
      mutate(race_factor = factor(race)) %>%
      within(race_factor <- relevel(race_factor, ref = 3)) %>%
      mutate(gender_factor = factor(sex, labels= c("Female","Male"))) %>%
      within(gender_factor <- relevel(gender_factor, ref = 2)) %>%
      mutate(score_factor = factor(score_text != "Low", labels = c("LowScore","HighScore")))
model <- glm(score_factor ~ gender_factor + age_factor + race_factor +
                            priors_count + crime_factor + two_year_recid, family="binomial", data=df)
summary(model)
```

**Northpointe allegation:**

The reverse logistic regression models are misspecified. And the relative risk ratios from the reverse regressions are miscalculated and misinterpreted.

**ProPublica Response:**

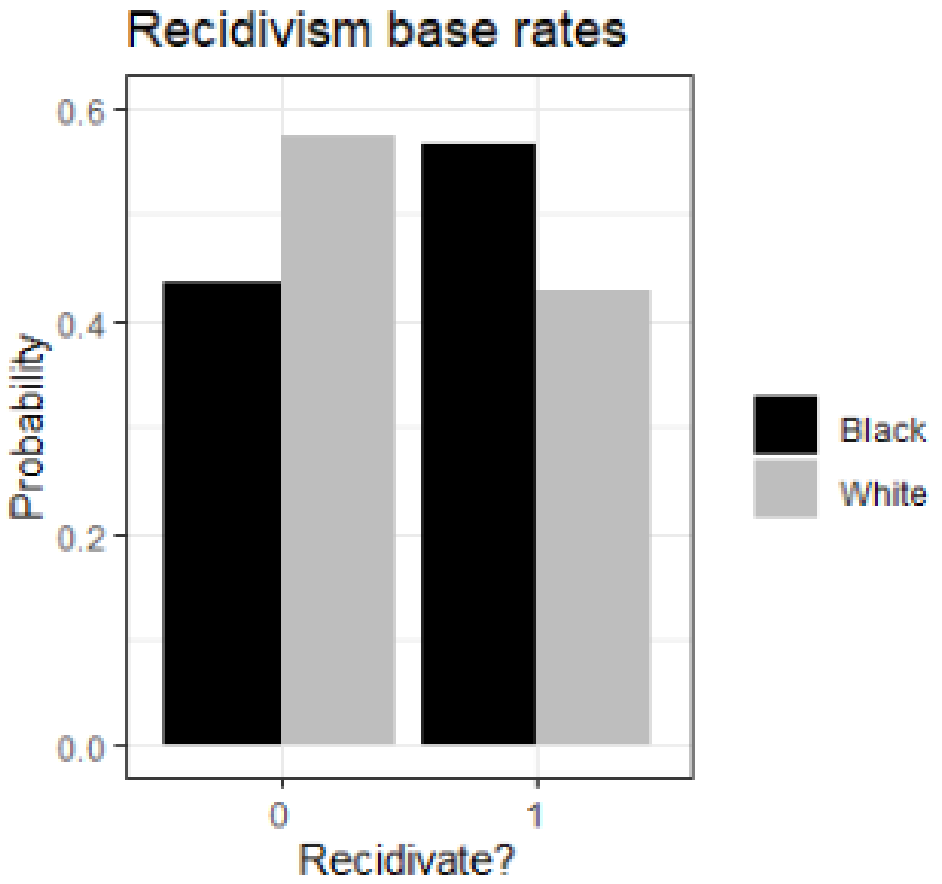
Our logistic model wasn't trying to predict who would recidivate. We were trying to identify a possible relationship between race and receiving a high score when controlling for other variables like age, gender and criminal history. We found that black defendants have greater odds of getting a high score that cannot be explained by these other factors. Then we even controlled for future recidivism, and still found that the racial gap couldn't be explained.

**Northpointe allegation:**

ProPublica neglected to consider the base rate in the interpretation of their results. This is an error in judgment about the probability of an event. The error occurs when information about the base rate of an event (e.g., low base rate of recidivism in a population) is ignored or not given enough weight.

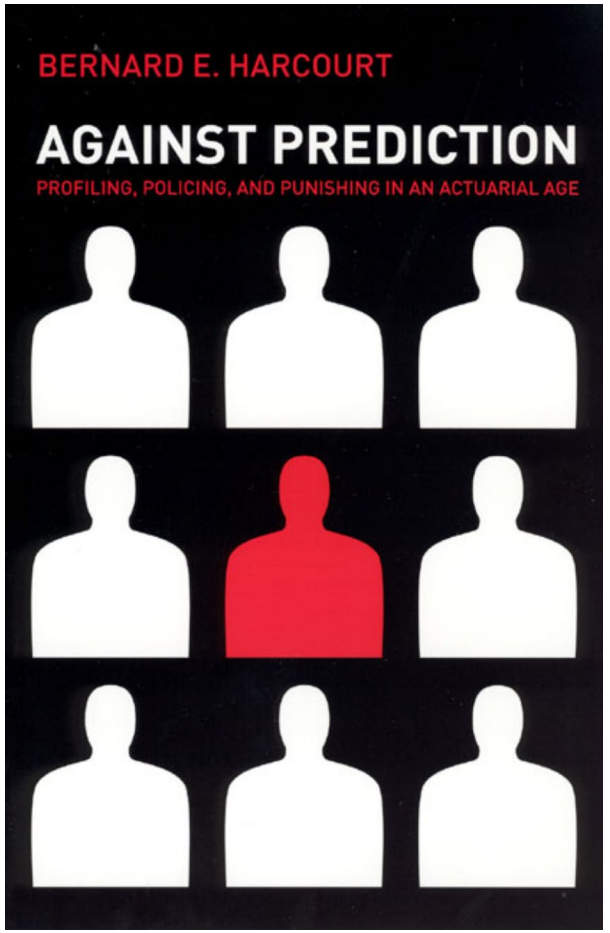
**ProPublica response:**

This is not correct. ProPublica accounted for the difference in recidivism rates in a statistical test known as a logistic regression. That test found that when adjusting for recidivism, criminal history, age and gender across races, black defendants were 45 percent more likely to get a higher score. In addition, we calculated likelihood ratios, which are useful for assessing how well a test performs independent of base rate. The

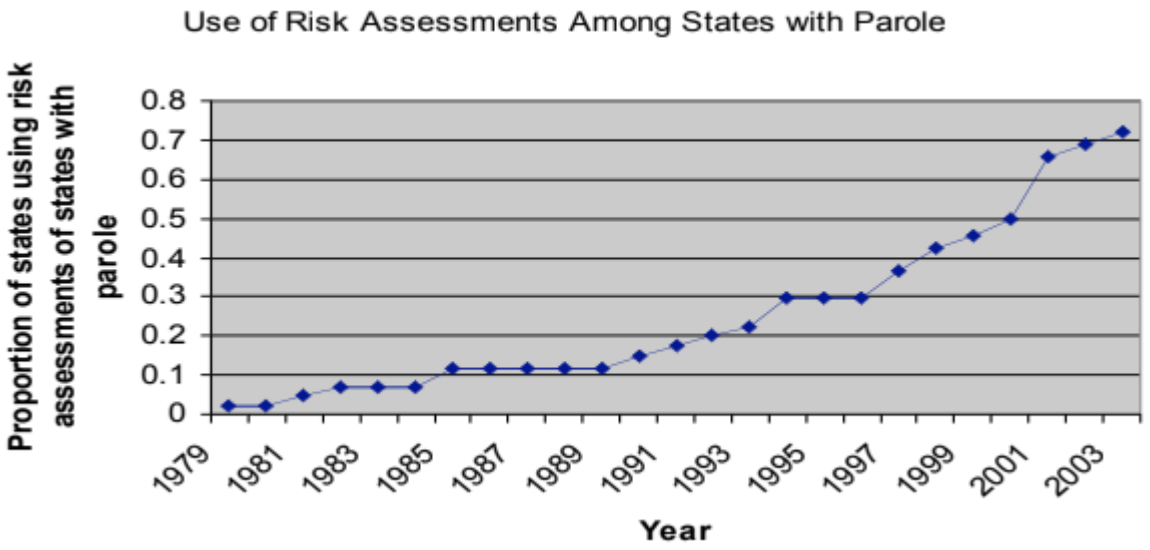


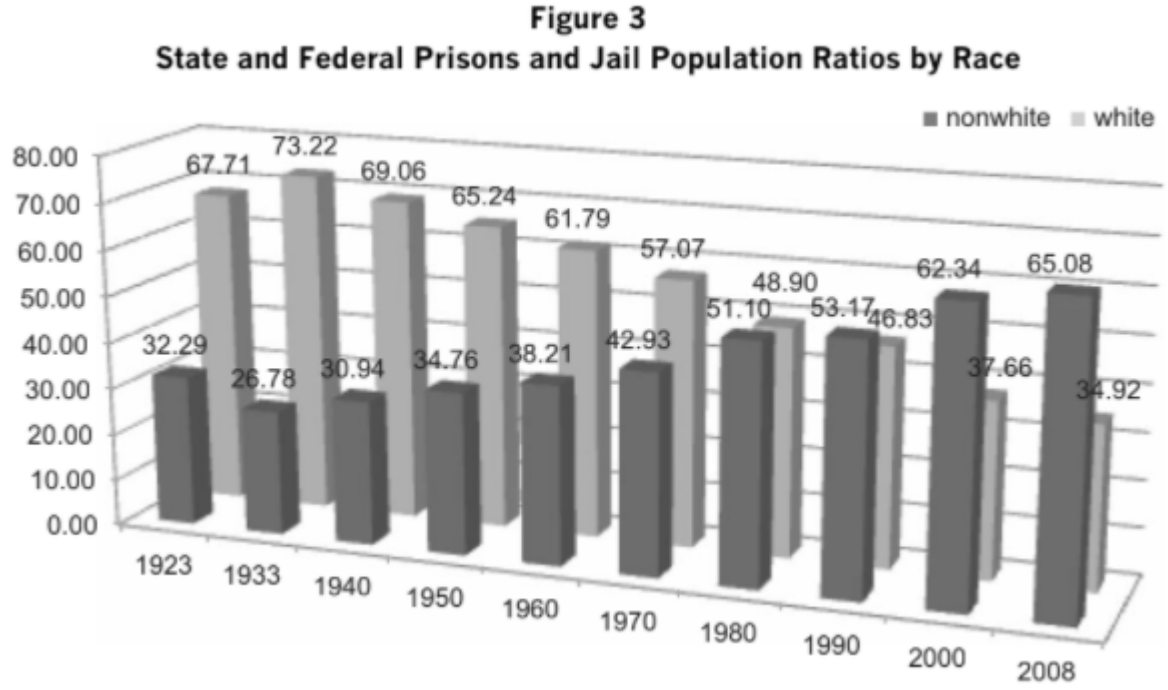
# What is bias in predictive modeling?

- Technical matters
- Ethical matters
- "*Unjustified* basis for differentiation"



# A case against prediction



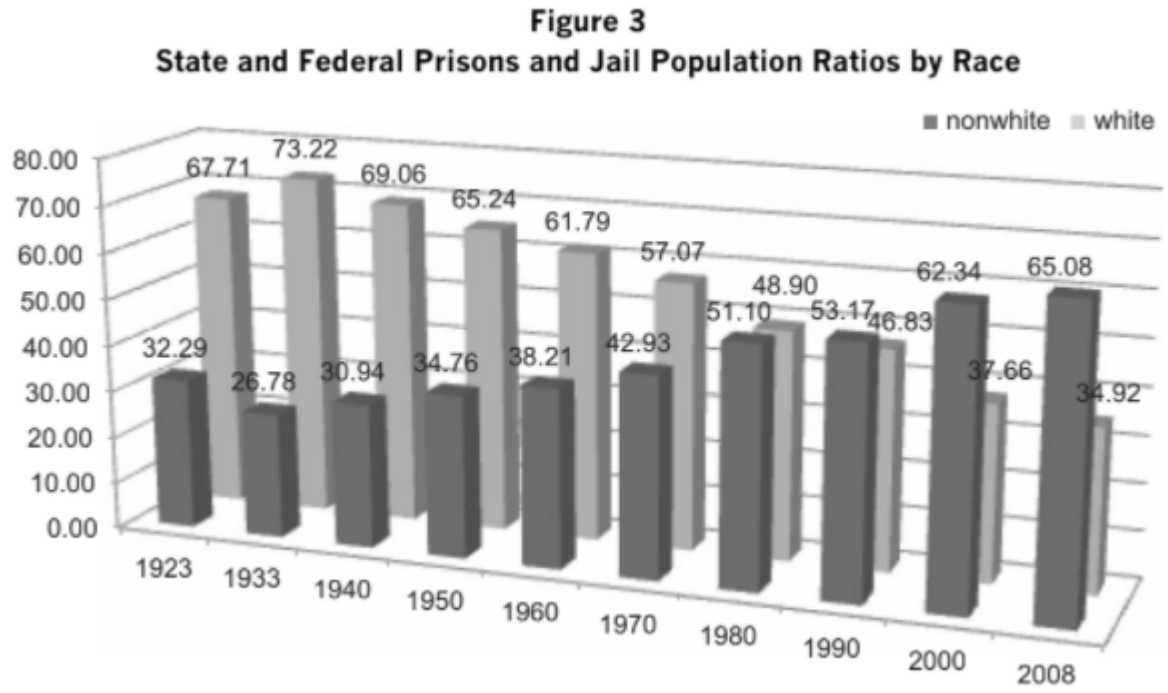



# A case against prediction

Racial disparity in justice system

***Over-sampling***

Predictive models assessing prior criminal records



# Sources of bias (in models)

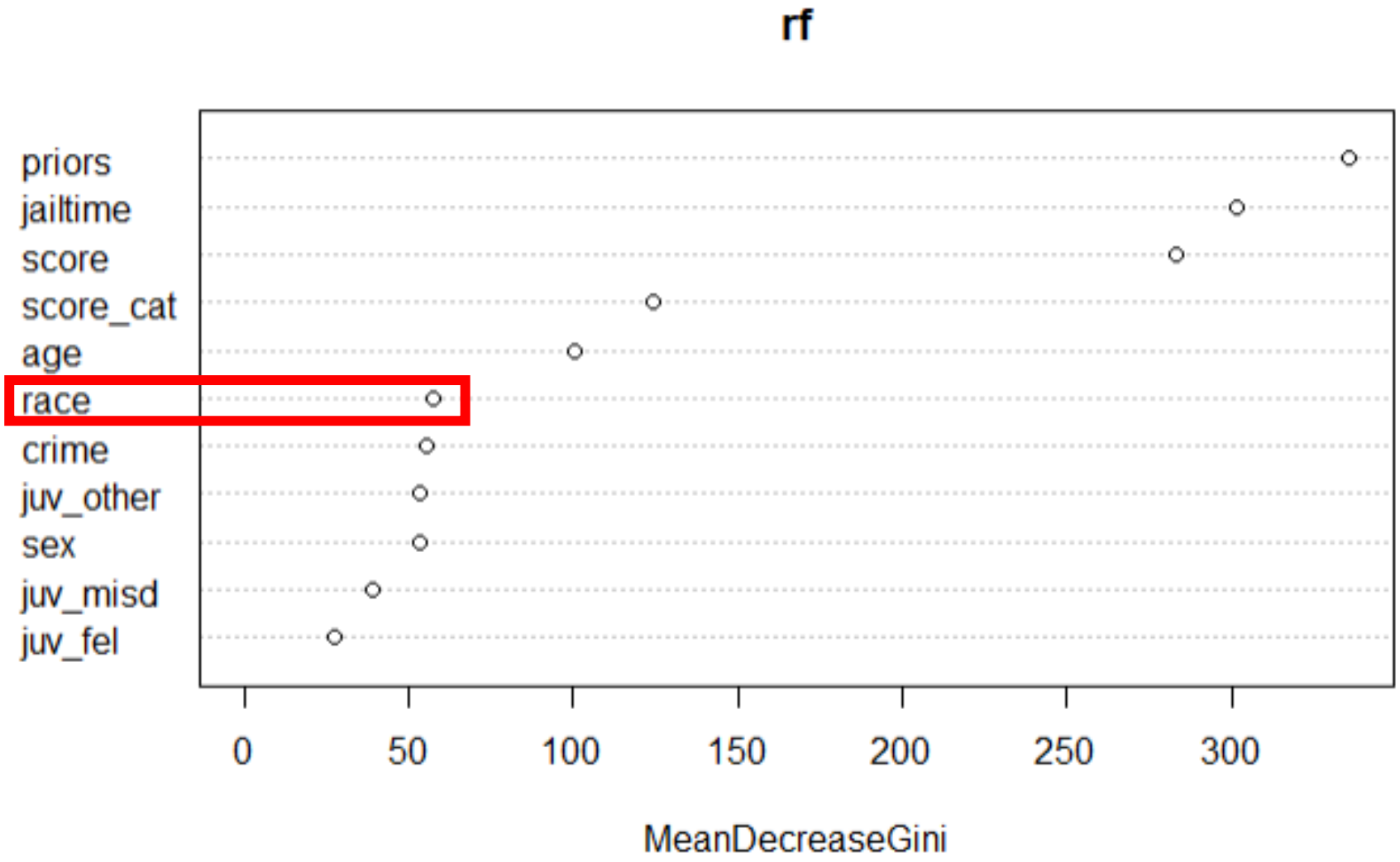
| | Example | Can we do anything? |
|---|---|---|
| Sampling bias | Participants in a study are all college students | Maybe (re-sampling) |
| Measurement bias | Survey asks loaded or poorly-worded questions | Maaaybe (transforming) |
| Systematic bias<br>Missing data | Over-policing leads to different base rates of recidivism | Probably not |
| Correlated/latent variables | Amount of prior arrests correlated with race | Maybe (transforming) |

# Criteria for fairness* in prediction

- Un-awareness (info on e.g. race/gender not used in prediction)
- Statistical parity (same probability of outcomes)
- Error ratio parity (same rates of prediction errors)

*Each come with problems, and solutions are currently debated

# Un-awareness



# Statistical parity – resampling on the outcome variable

| Race | Recidivate? |
|---|---|
| W | 0 |
| W | 0 |
| W | 0 |
| W | 1 |
| B | 0 |
| B | 0 |
| B | 1 |
| B | 1 |

**25%**

**50%**

**Resample**

| Race | Recidivate? |
|---|---|
| W | 0 |
| W | 1 |
| W | 0 |
| W | 1 |
| B | 0 |
| B | 1 |
| B | 0 |
| B | 1 |

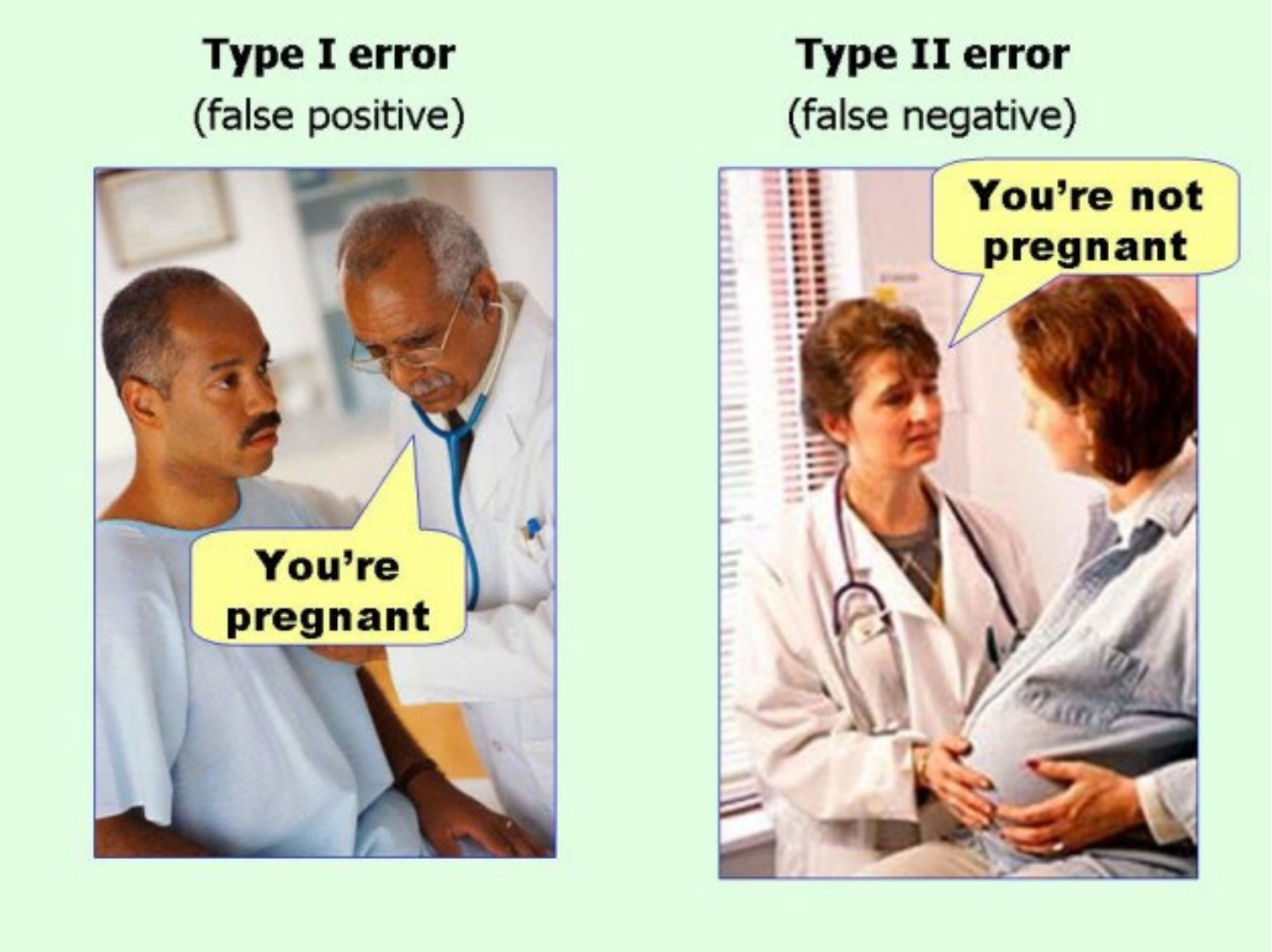
**50%**

**50%**

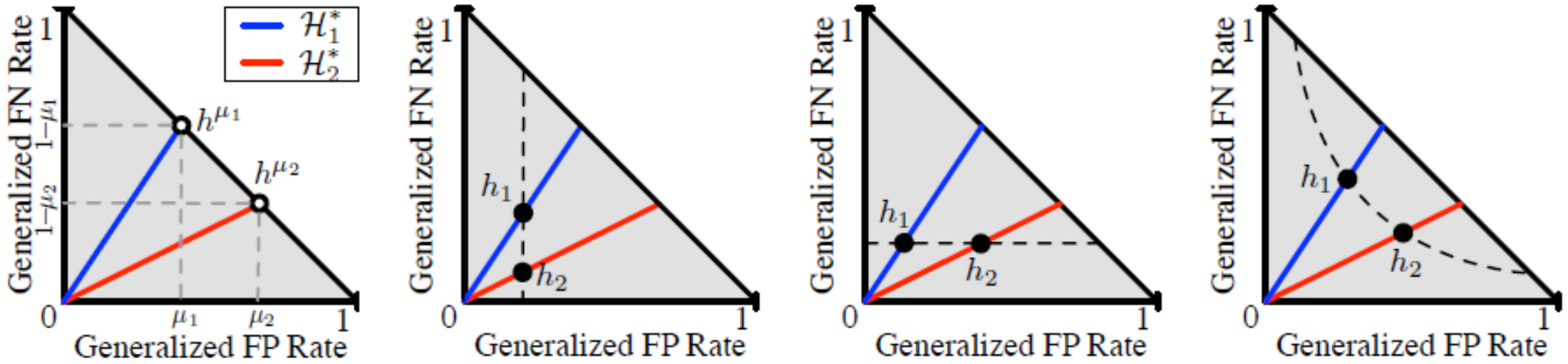
# Error ratio parity

The model is **wrong** at some rate…

**Types** of errors are not the same



# Error ratio parity – impossible to optimize fairly?




(a) Possible cal. classifiers $\mathcal{H}_1^*, \mathcal{H}_2^*$ (blue/red).

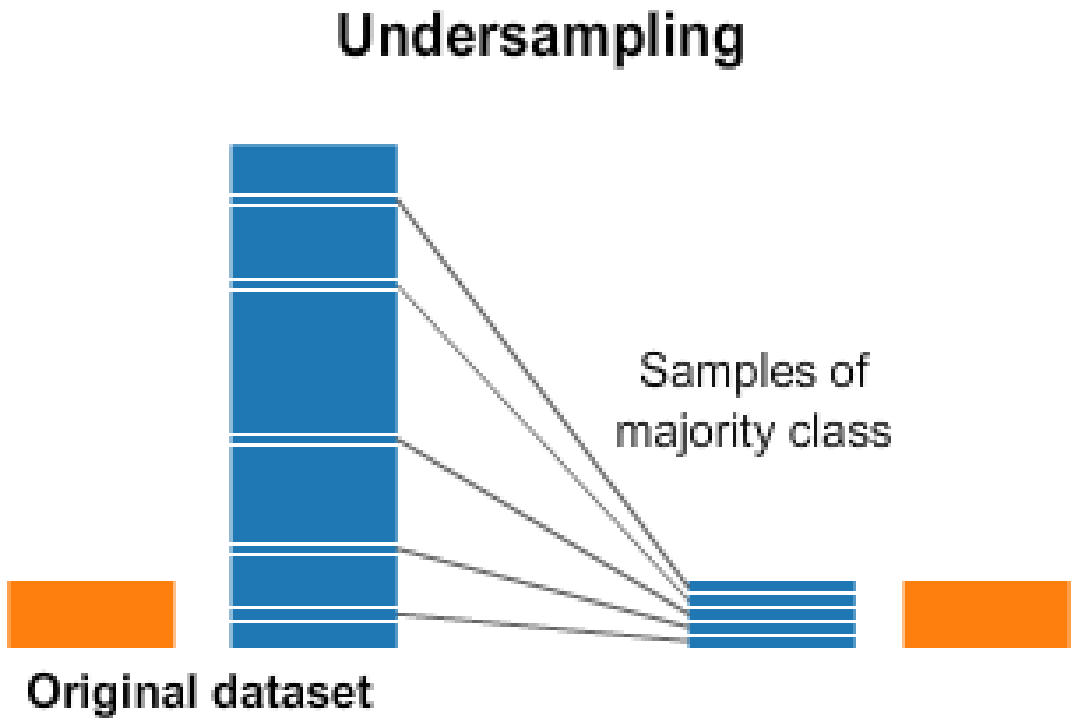
(b) Satisfying cal. and equal F.P. rates.

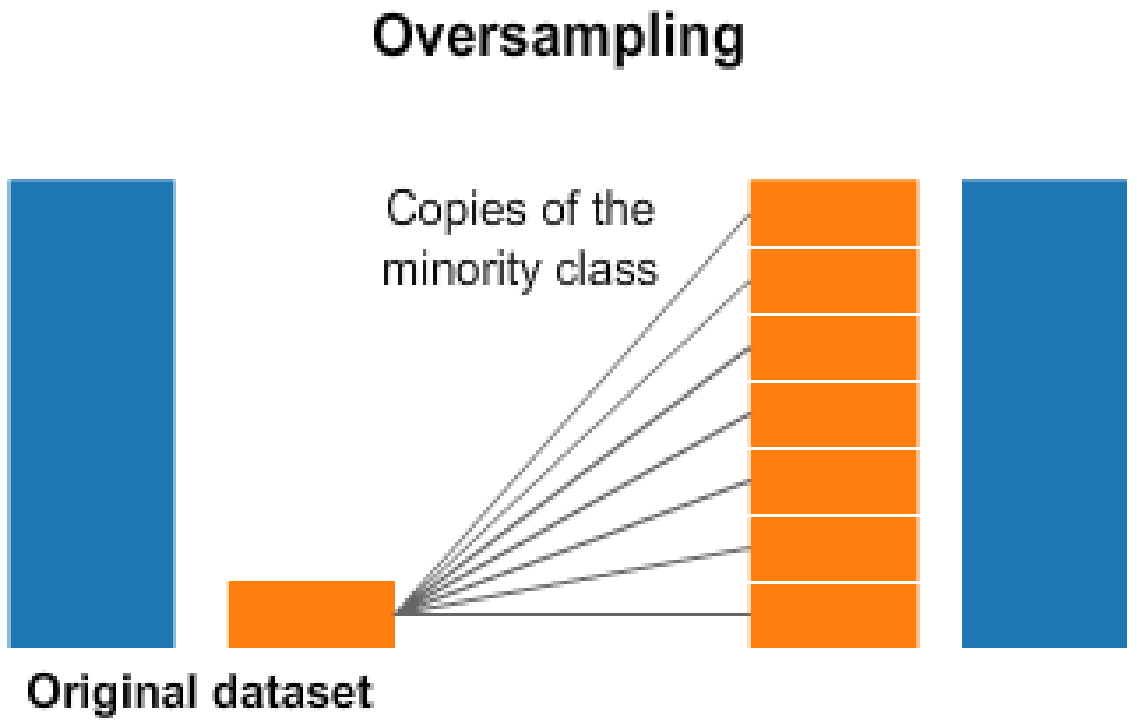
(c) Satisfying cal. and equal F.N. rates.

(d) Satisfying cal. and a general constraint.

Figure 1: Calibration, trivial classifiers, and equal-cost constraints – plotted in the false-pos./false-neg. plane. $\mathcal{H}_1^*, \mathcal{H}_2^*$ are the set of cal. classifiers for the two groups, and $h^{\mu_1}, h^{\mu_2}$ are trivial classifiers.

# A tool we can use: re-sampling



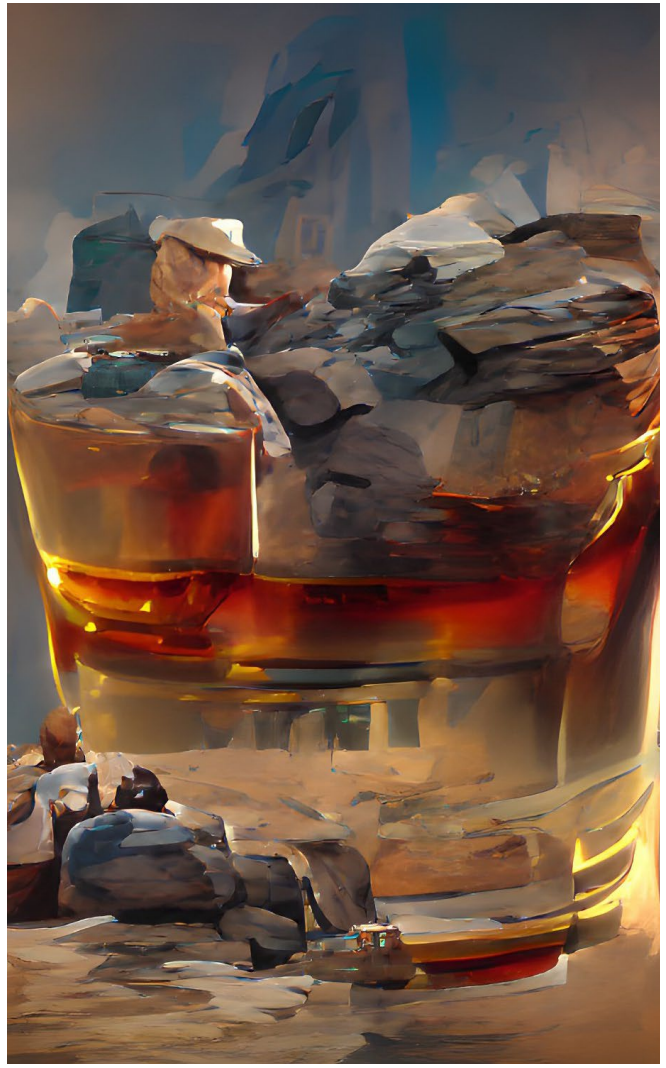



# Neural networks

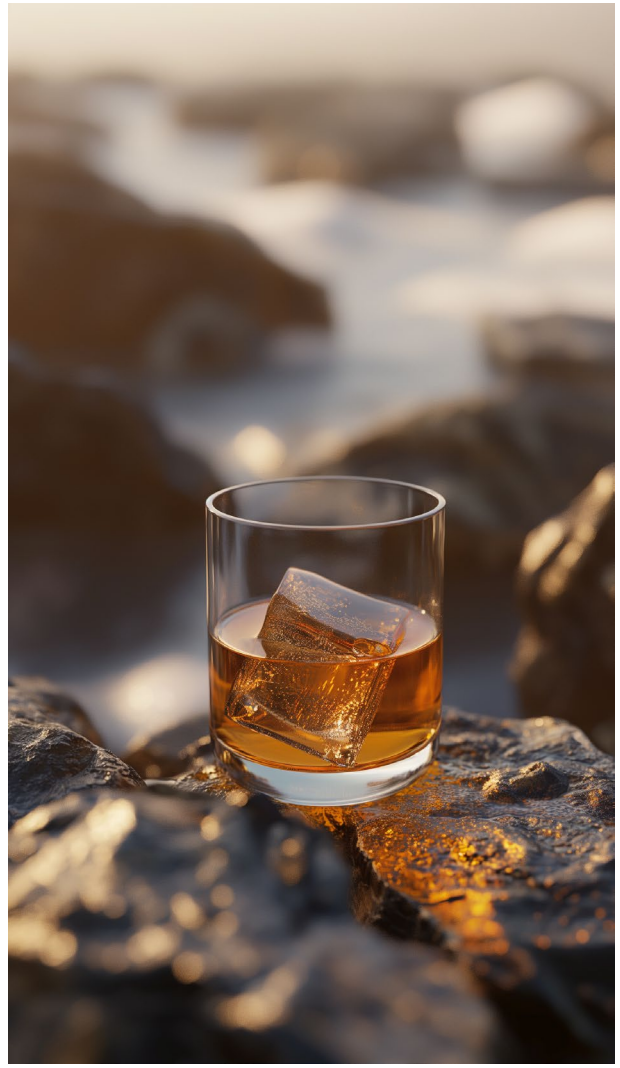
The **state of the art** for predictive analytics

- Steep learning curve
- Require dedicated packages
- Difficult to interpret
- **Extremely powerful** predictive tools

“Whisky on the rocks” 2022
wombo.art



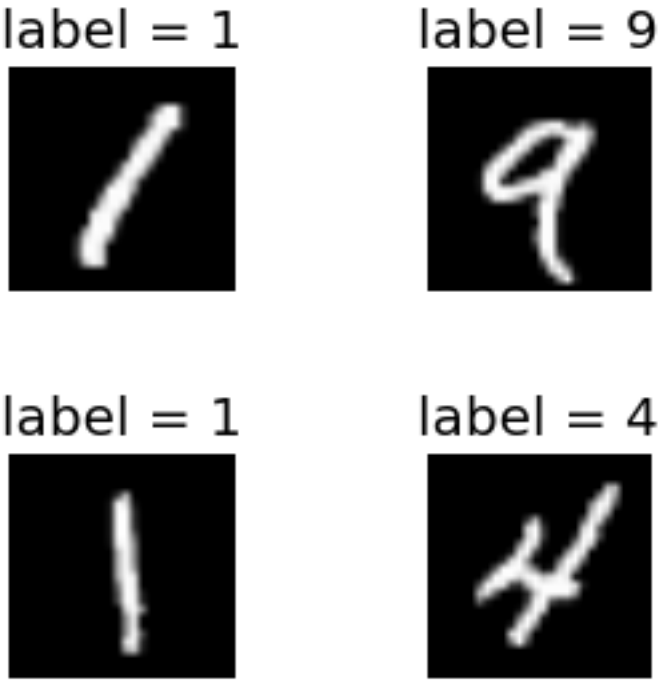
“Whisky on the rocks” 2024
Midjourney



# Neural networks are supervised models (mostly)

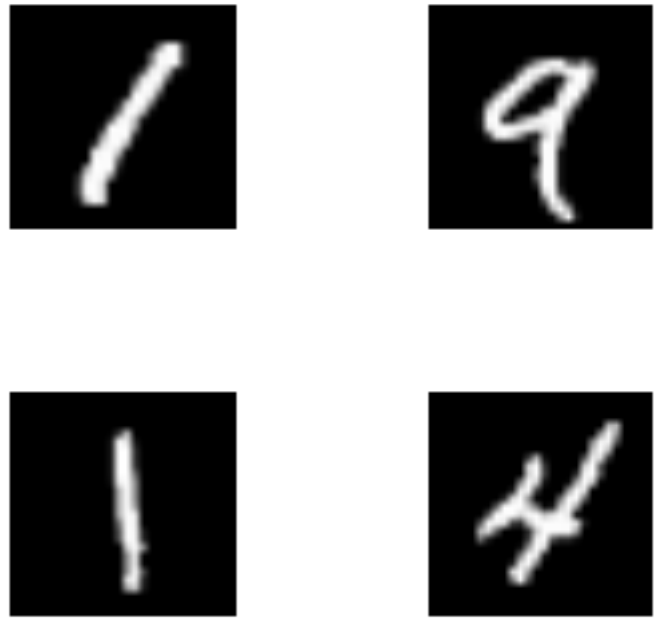
**Supervised**

Ground truth available

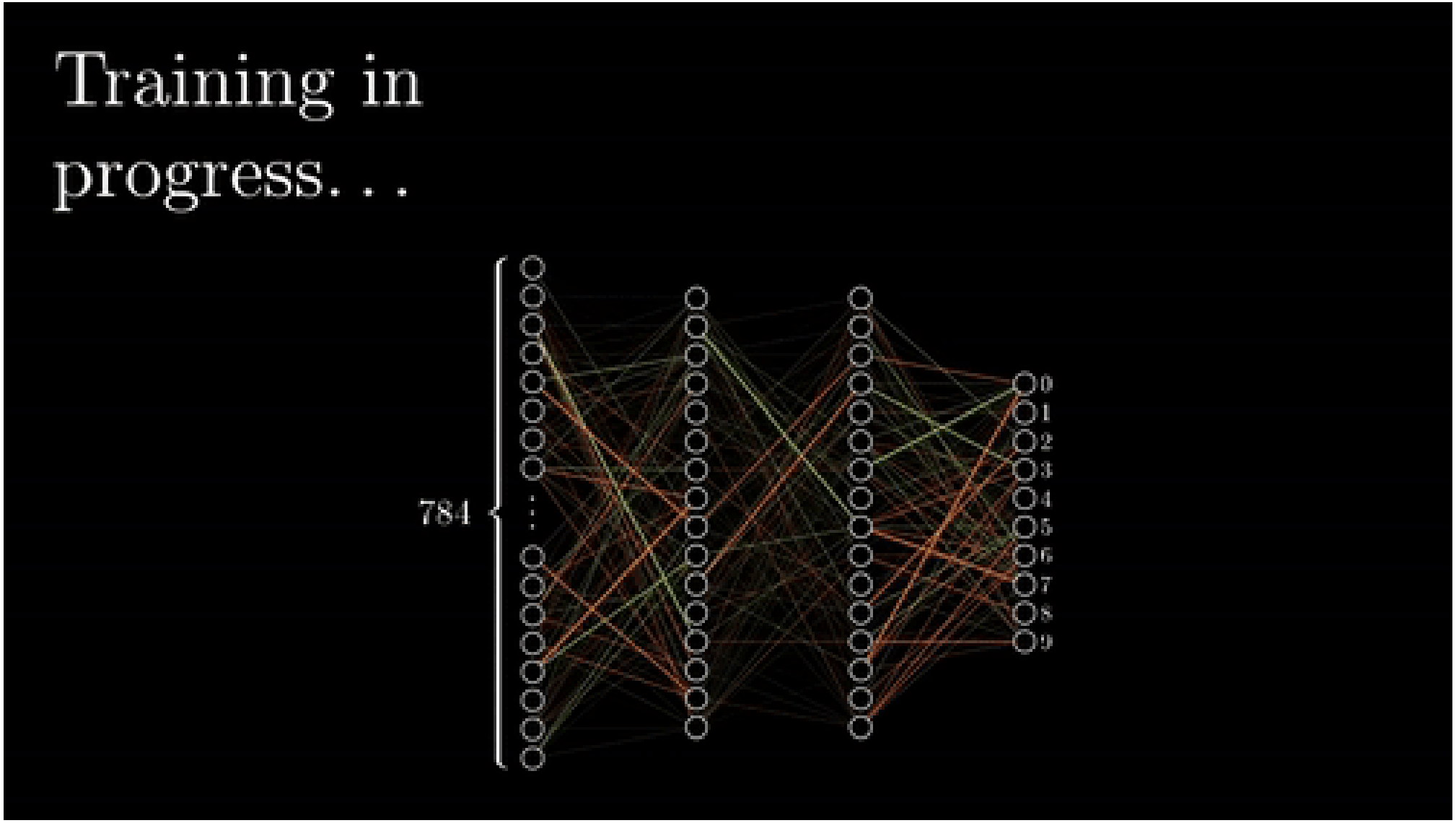



**Unsupervised**

Ground truth unavailable

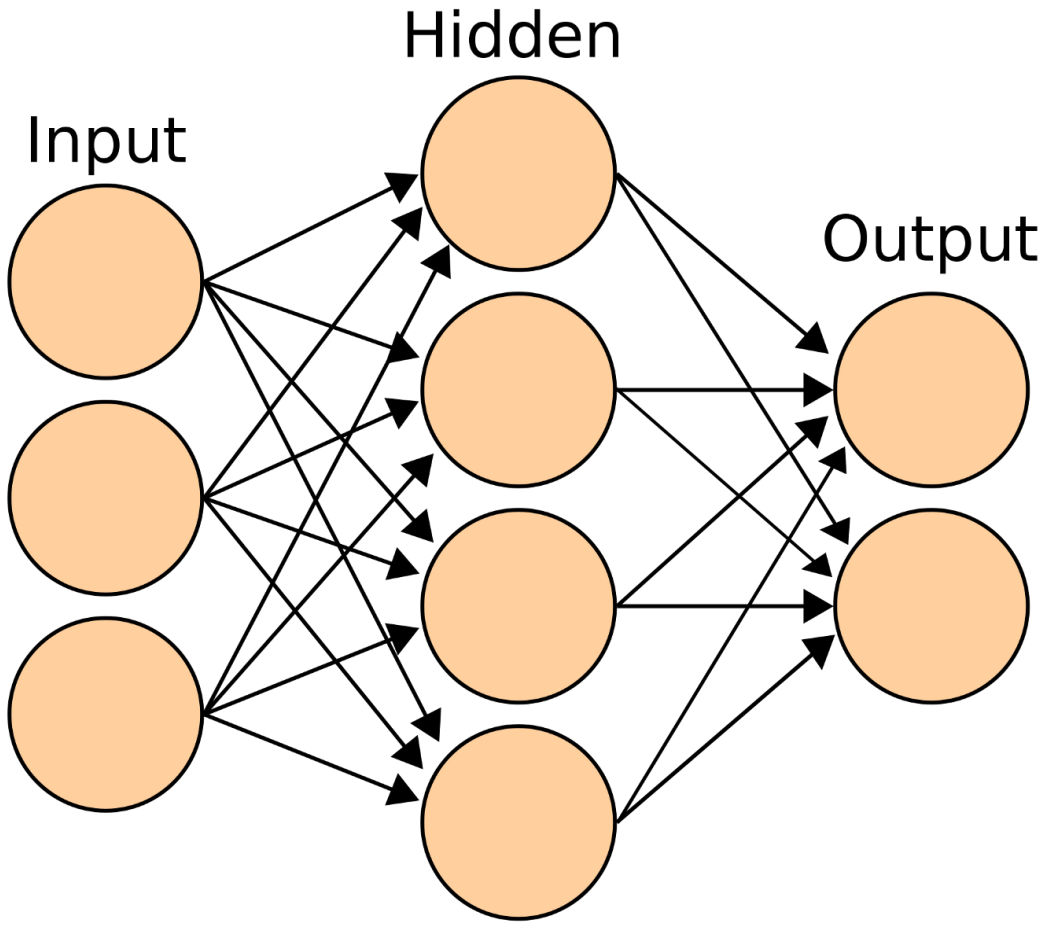


# Artificial neural network (ANN) model



# Artificial neural network (ANN) model

| Education_Level | Dependents | Prison_Offense |
|---|---|---|
| At least some college | 3 or more | Drug |
| Less than HS diploma | 1 | Violent/Non-Sex |
| At least some college | 3 or more | Drug |
| Less than HS diploma | 1 | Property |
| Less than HS diploma | 3 or more | Violent/Non-Sex |
| Less than HS diploma | 2 | |



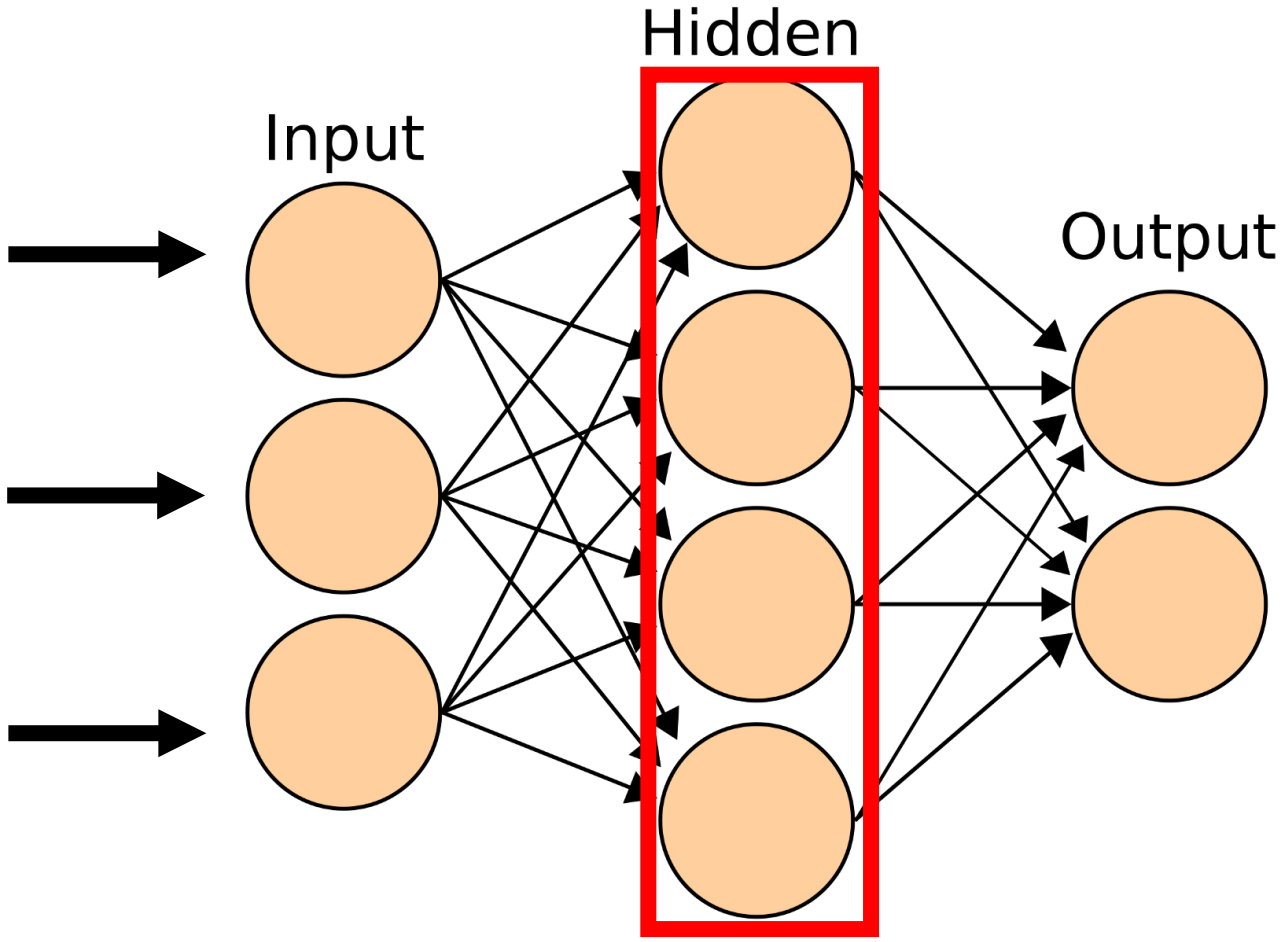

Did recidivate?

Yes

No

# Artificial neural network (ANN) model

| Education_Level | Dependents | Prison_Offense |
|---|---|---|
| At least some college | 3 or more | Drug |
| Less than HS diploma | 1 | Violent/Non-Sex |
| At least some college | 3 or more | Drug |
| Less than HS diploma | 1 | Property |
| Less than HS diploma | 3 or more | Violent/Non-Sex |
| Less than HS diploma | 2 | |

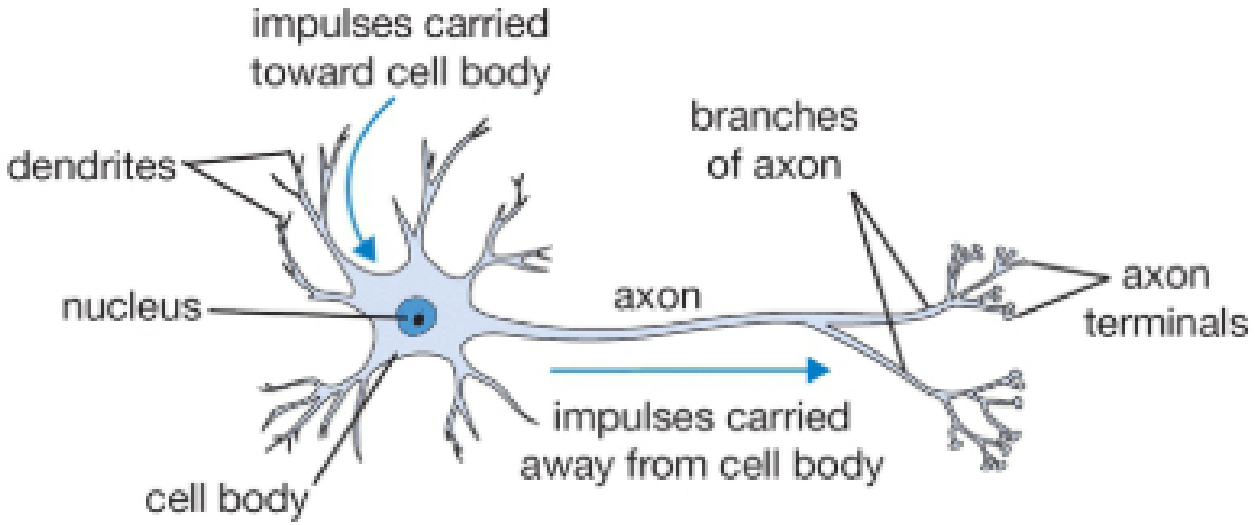


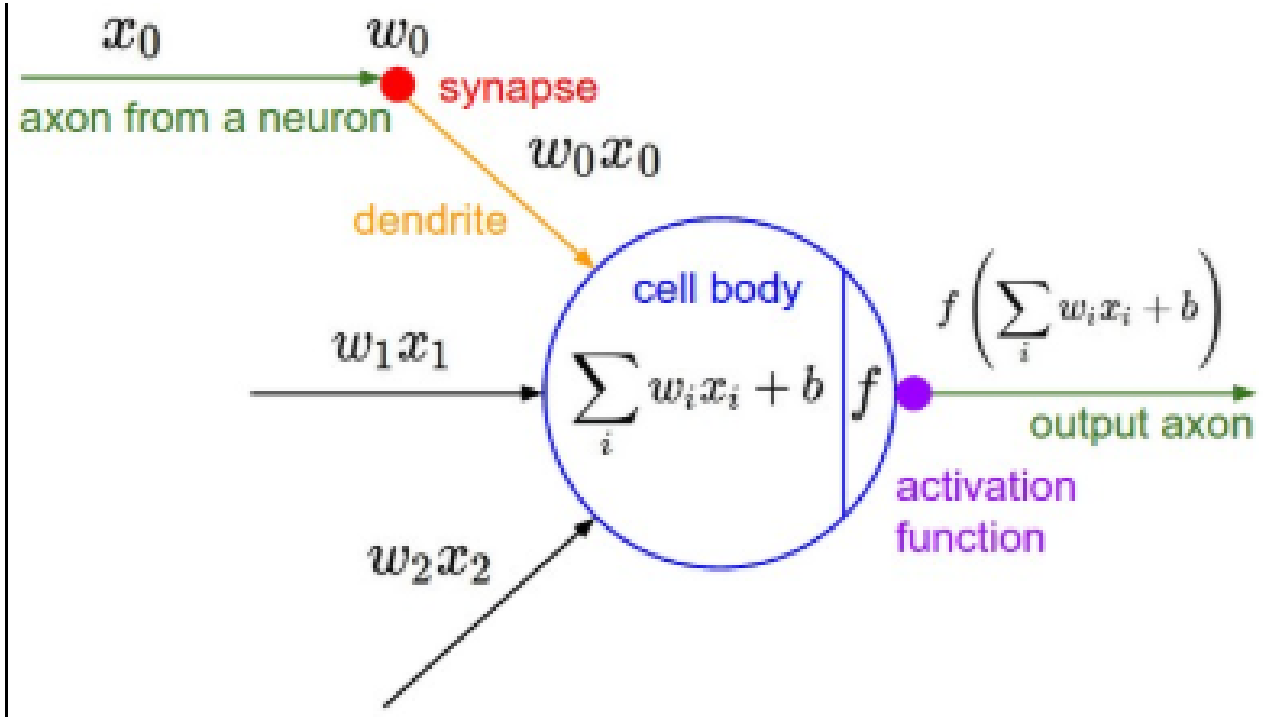

Did recidivate?

Yes

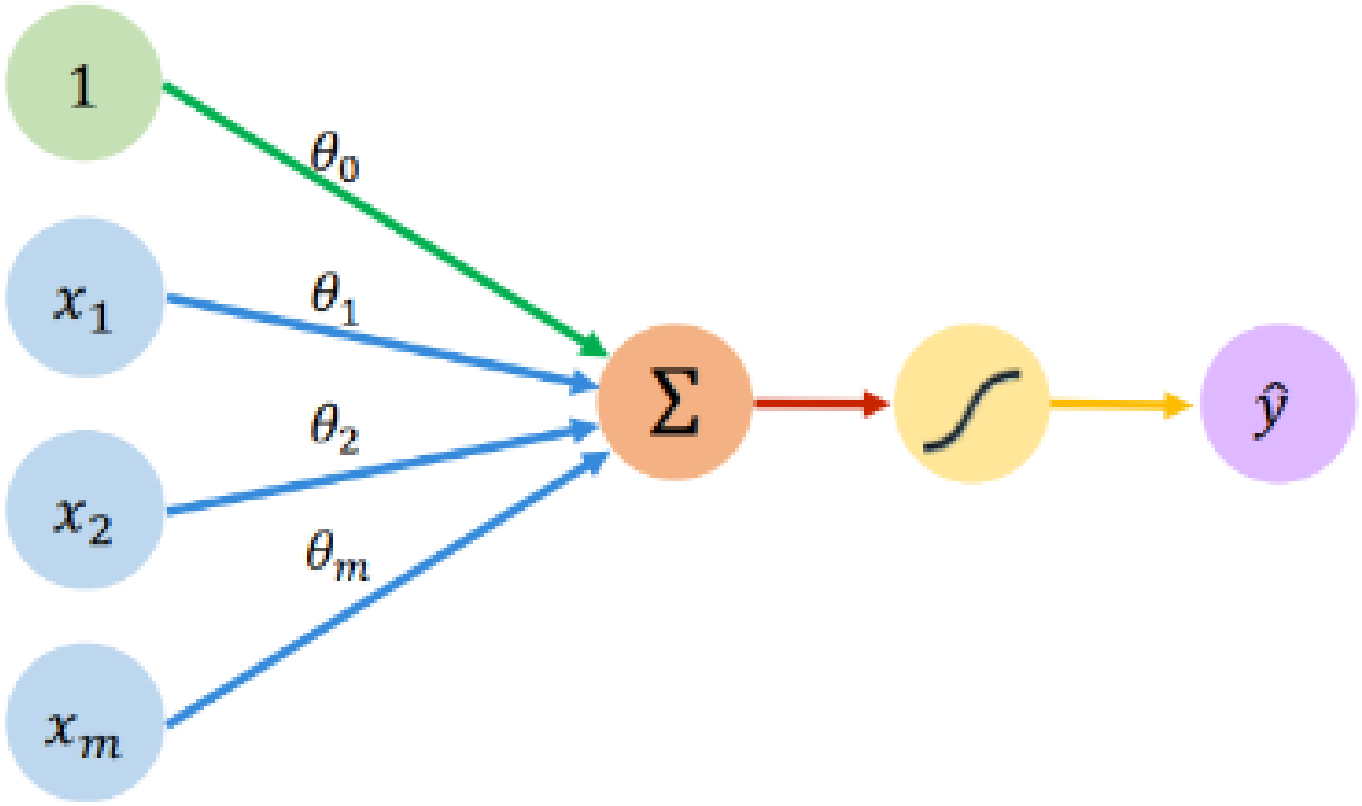
No

# Artificial neural network (ANN) model



# Artificial neural network (ANN) model




Inputs Weights Sum Non-Linearity Output

Output

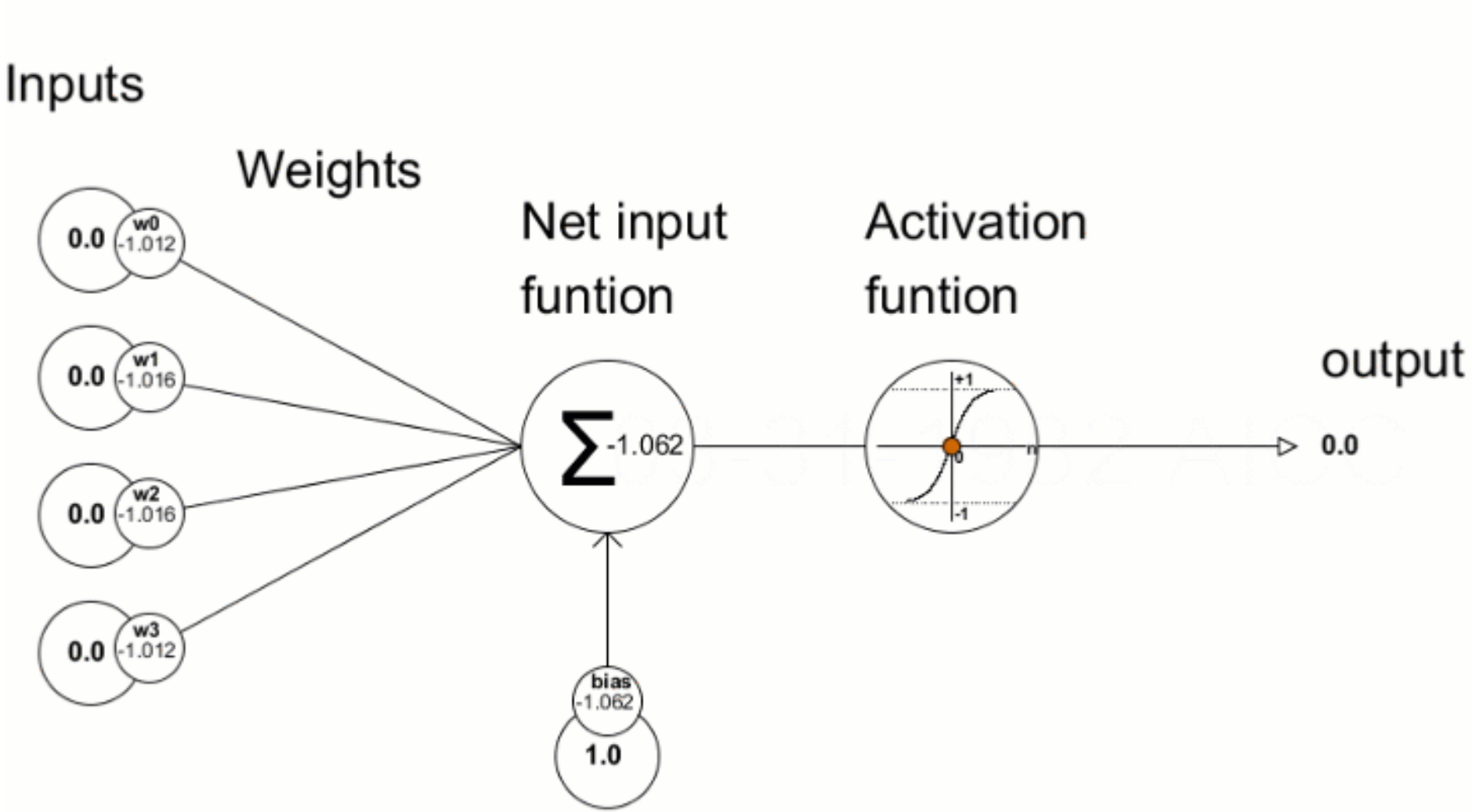
Linear combination of inputs

$$\hat{y} = g\left(\theta_0 + \sum_{i=1}^{m} x_i\ \theta_i\right)$$

Non-linear activation function

Bias

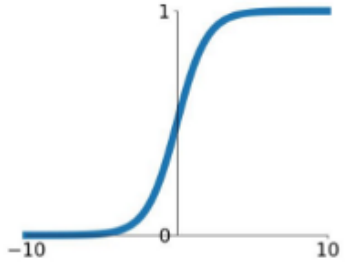
# What makes a neuron fire – **activation functions**



# What makes a neuron fire – activation functions

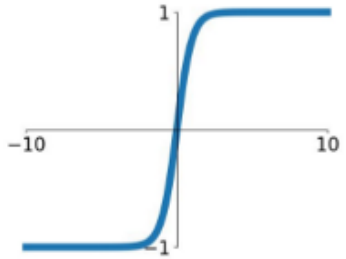
**Sigmoid**

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



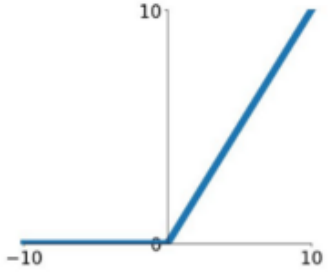

**tanh**

$$\tanh(x)$$



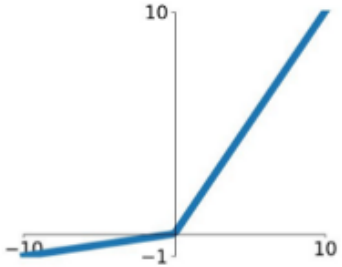

**ReLU**

$$\max(0, x)$$




**Leaky ReLU**

$$\max(0.1x, x)$$

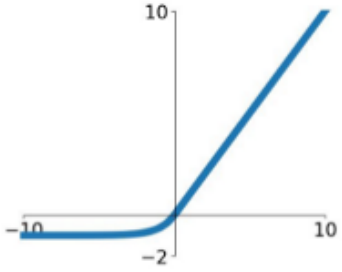



**Maxout**

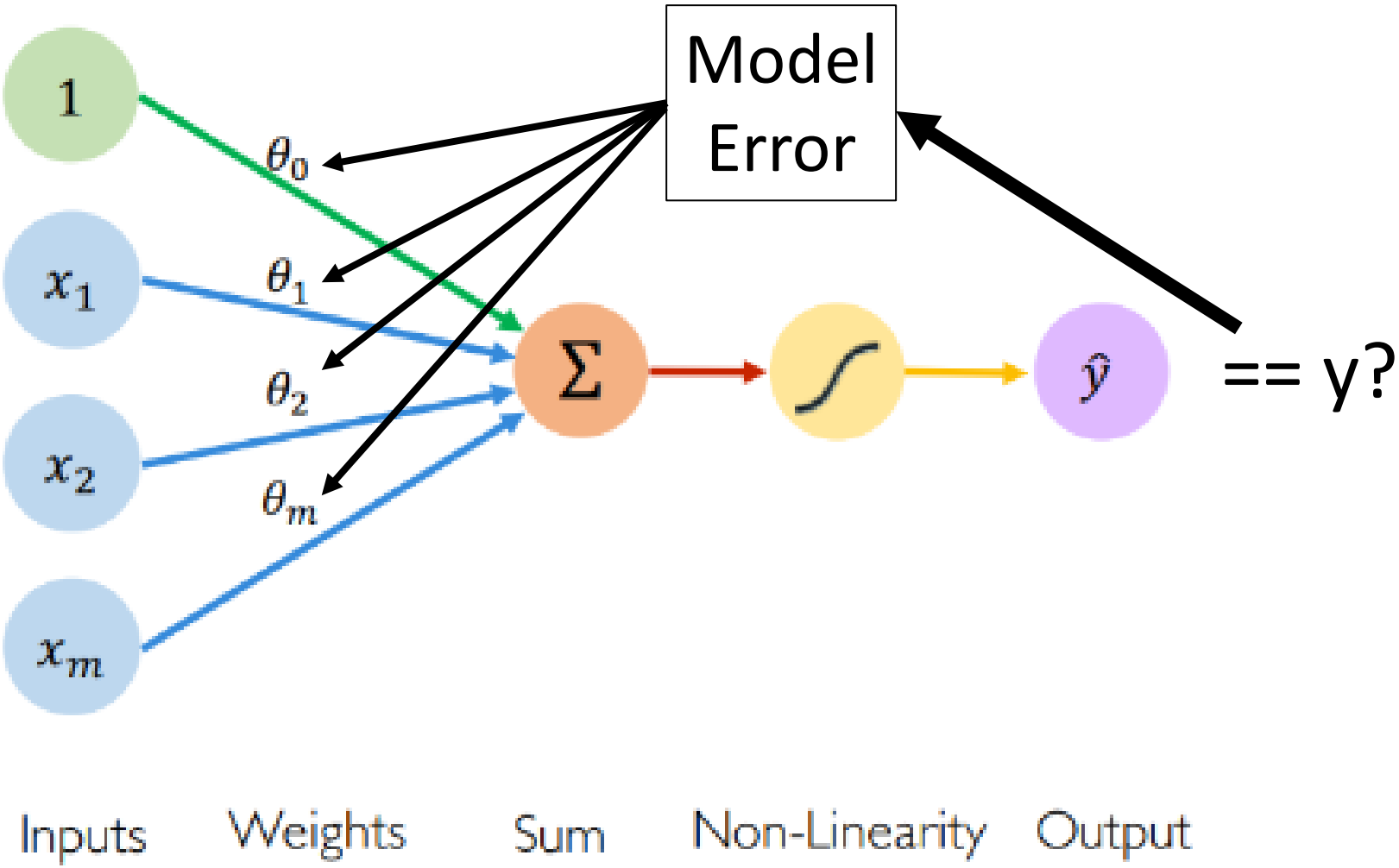
$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

**ELU**

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

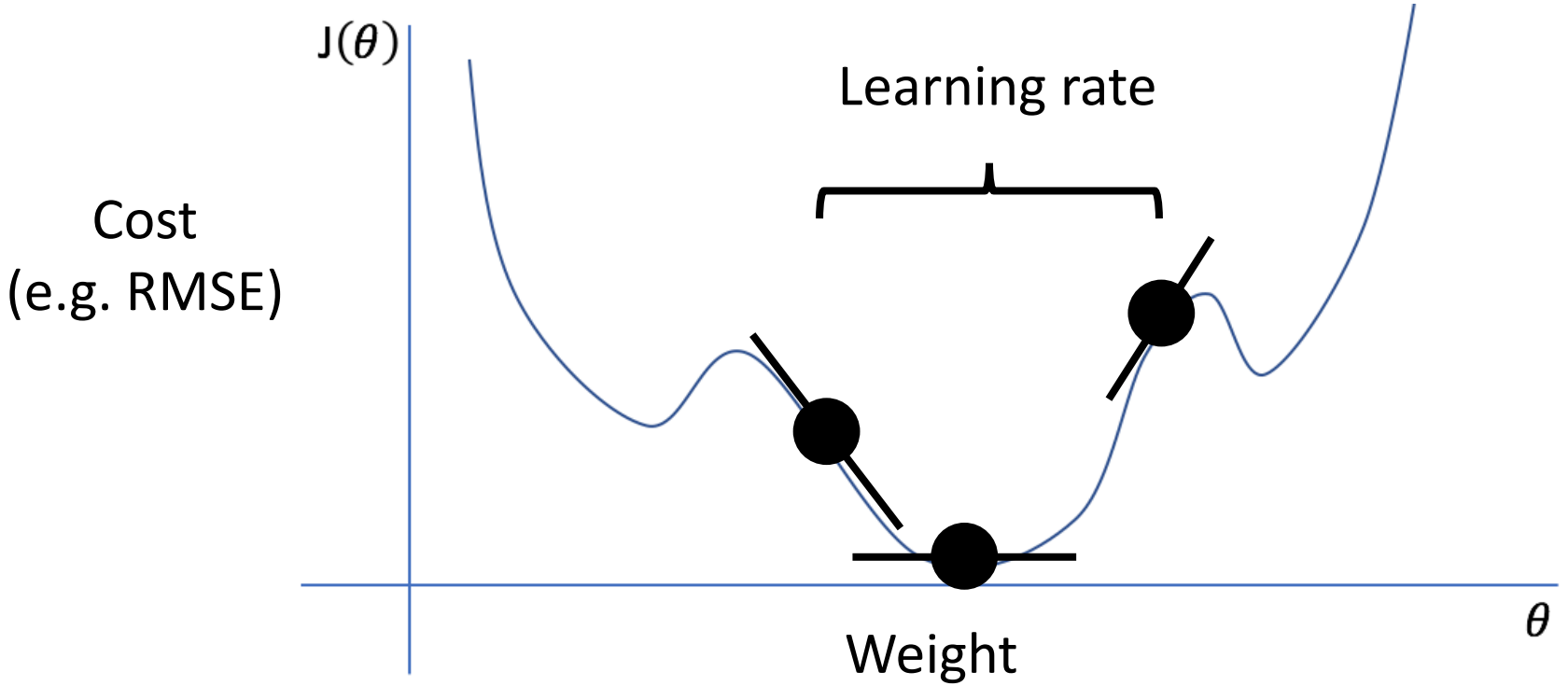


# Learning from mistakes - **backpropagation**

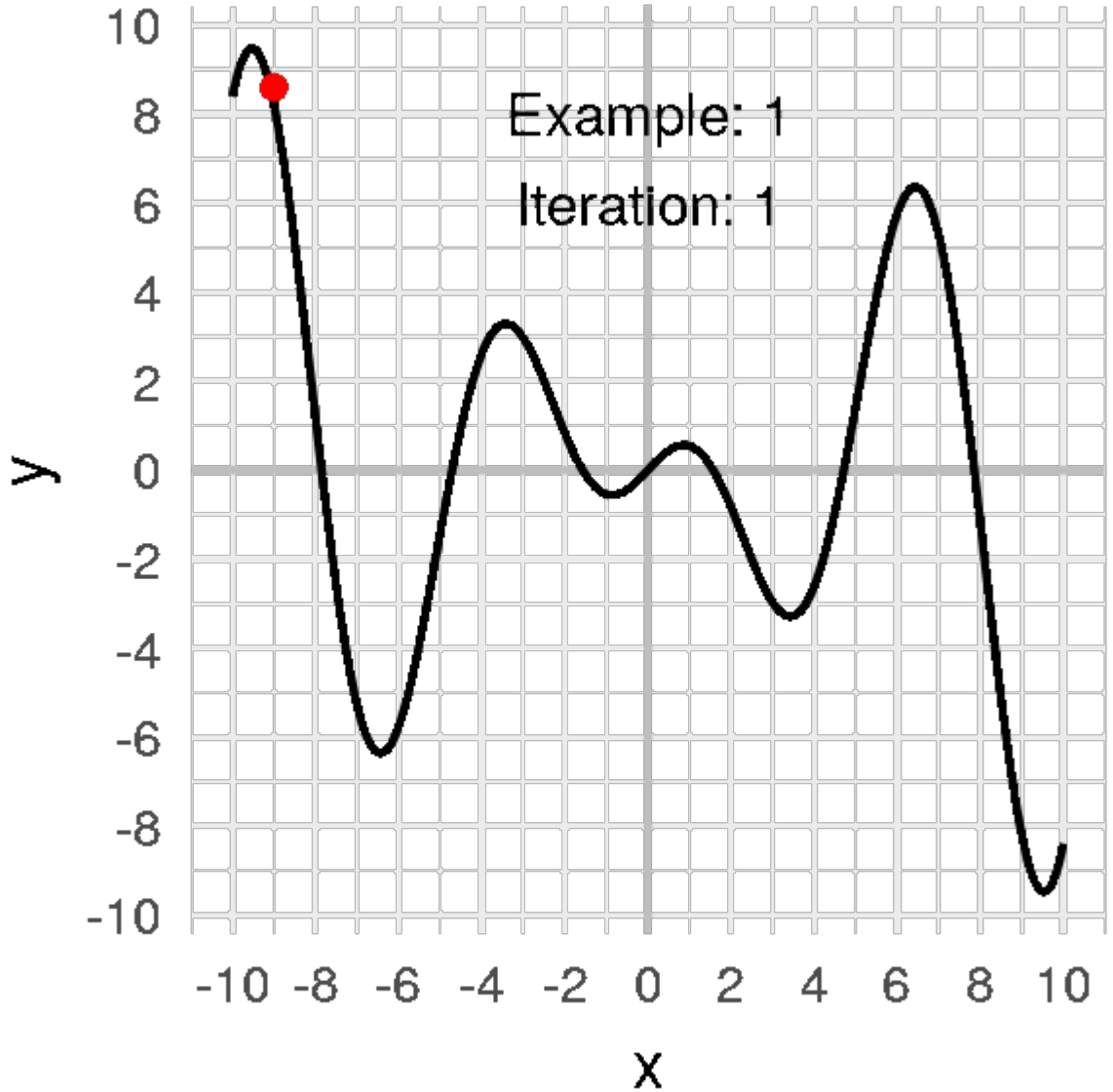


# How to adjust weights – cost function

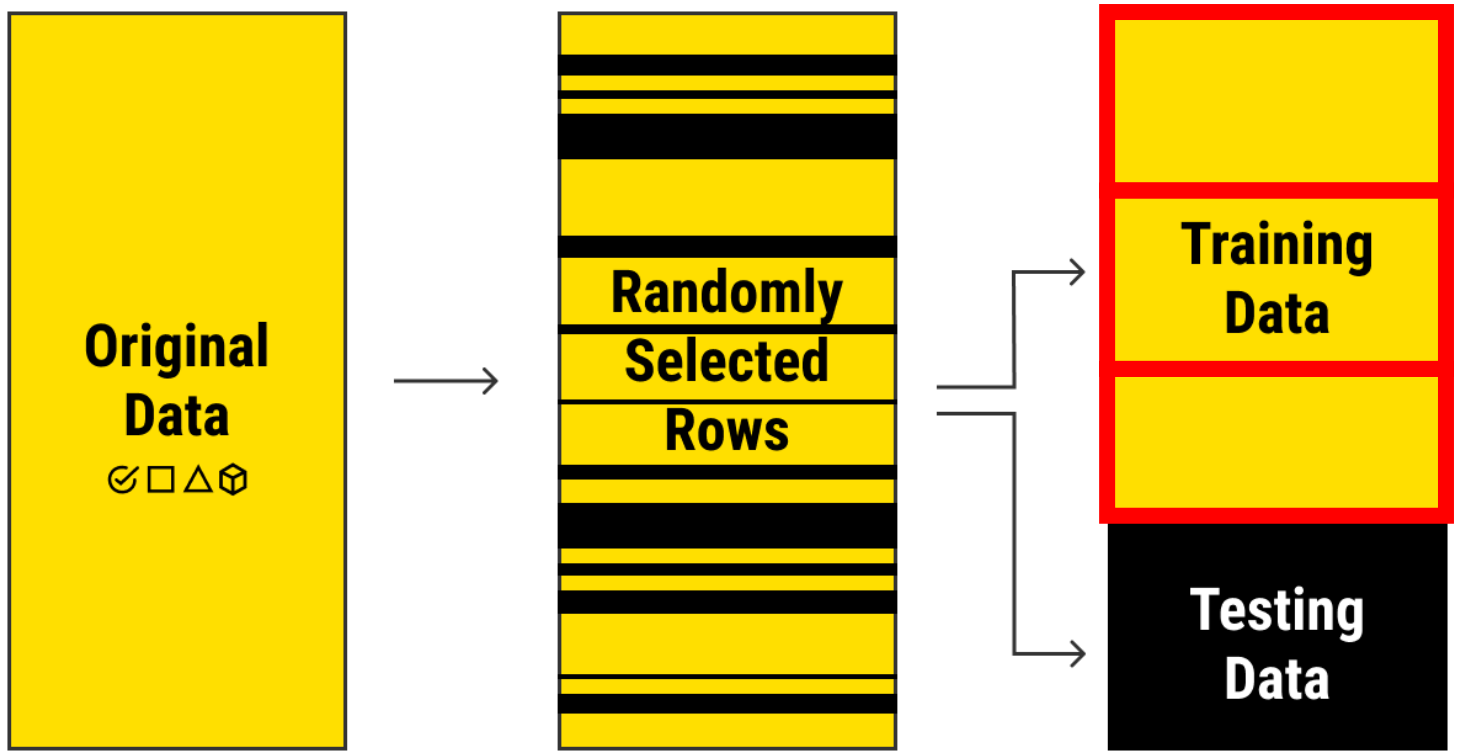
- The process of adjusting weights IS the learning/fitting process
- The **cost function** tells you which direction to adjust them
- We don't know the cost function



# How to adjust weights – **cost function**



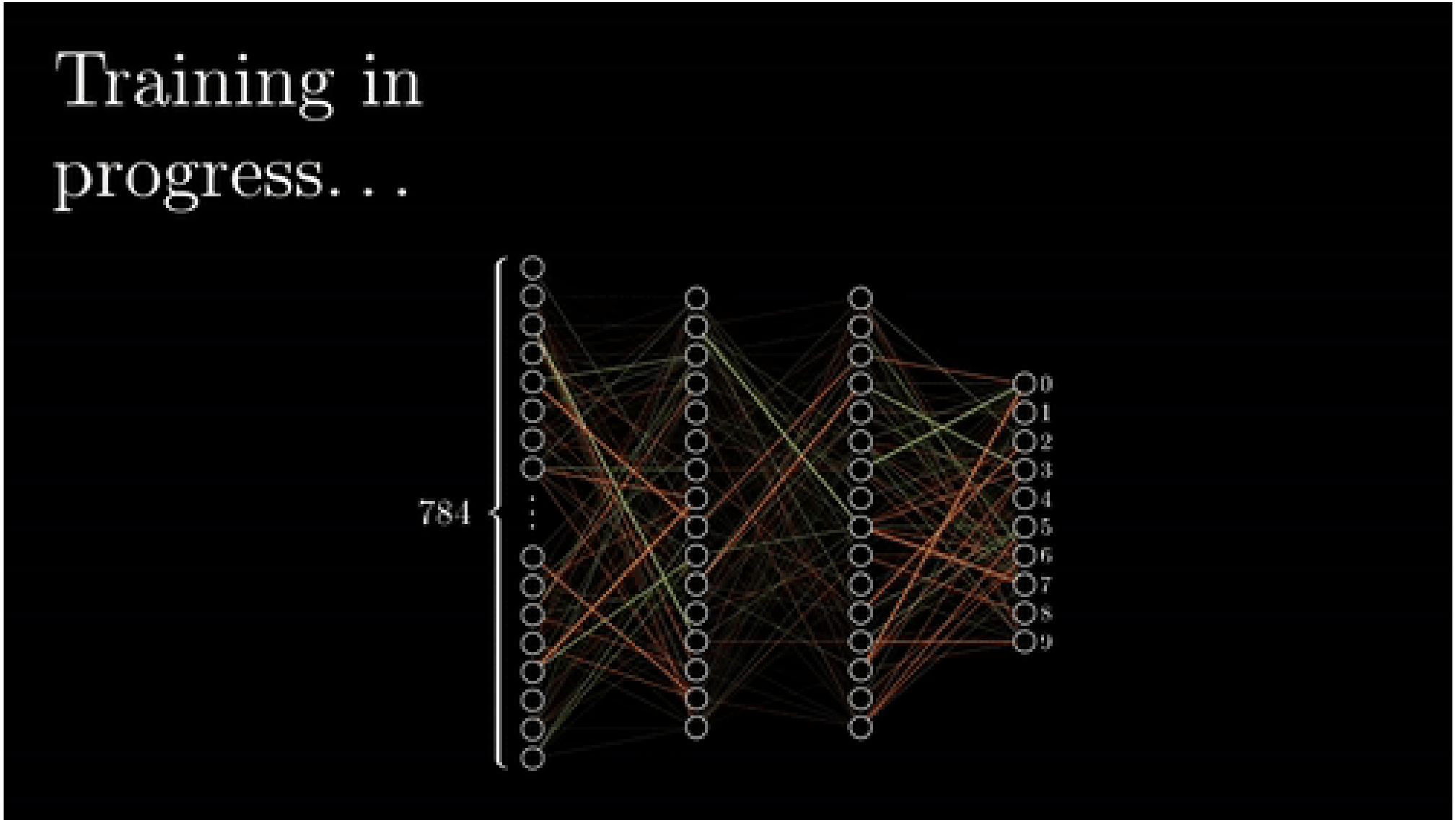
# How to assess cost function – cross-validation




Splitting original data:

- **Training**
  - Training data
  - Validation data
- **Testing**

# Minimize the cost function



# Endless network architectures