CS35L Assignment 10

Junhong Wang

1. Review

1.1 Article

Amazon face-detection technology shows gender and racial bias, researchers say

Protesters hold images of Amazon CEO at Amazon Headquarter
and complain its facial recognition system (Oct. 31, 2018, in Seattle.)

1.2 Summary

Amazon's facial detection technology often misidentifies women, especially those with darker skin. This is a serious problems because it implies gender and racial discriminations. The researchers performed some tests. The table below summarises the result.

Subject	Error Rate (misclassified as opposite sex)
Darker-skinned women	31%
Lighter-skinned women	7%
Darker-skinned men	1%
lighter-skinned men	0%

AI can learn biases from its creators, which absolutely should be avoided. Biased facial detection technology has potential to be abused and threatens privacy and civil liberties.

1.3 Main Idea

- AI can learn biased from human creators
- Biased AI can be a threat to civil liberties

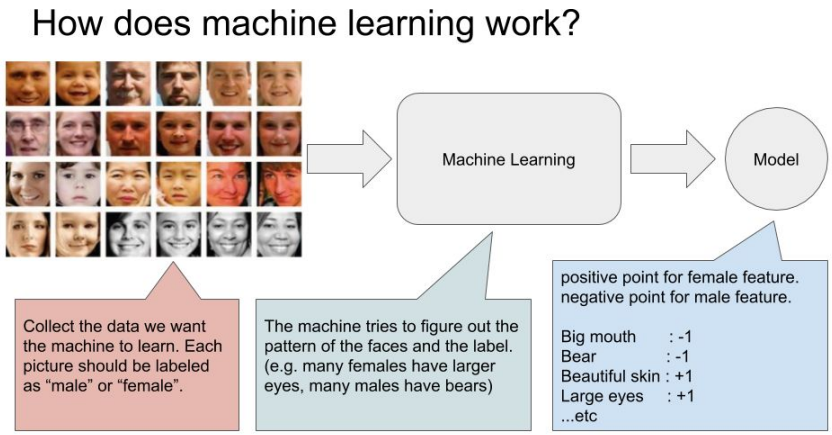
1.4 My Reaction

I believe biases in technology is a serious problem. It is very sad that creators of AI unintentionally add biases to the system. I wonder what exactly is making AI biased and how to solve the problem. My question is what do people really mean by AI is biased.

2. Introduction

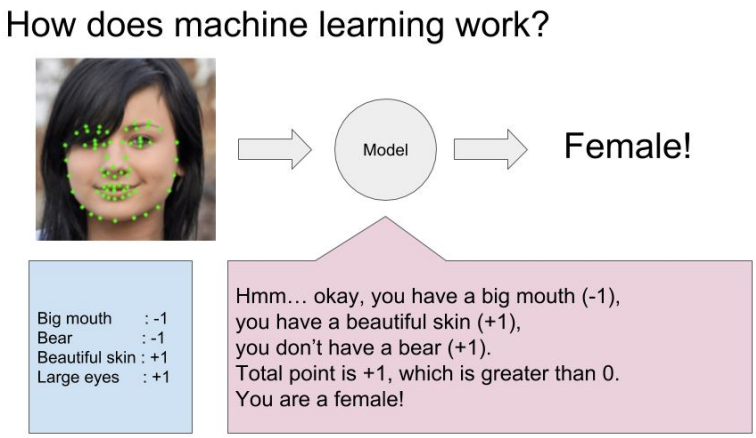
2.1 Training Phase

To understand how AI learn bias, we need to first understand how facial recognition, or more broadly speaking, how machine learning works. In general, we give the machine bunch of data. Then the machine will look for patterns and “recognize”, for example, that men’s faces have certain features that women’s don’t have, and vice versa. Finally, the machine will store all these information into a model.



2.2 Testing Phase

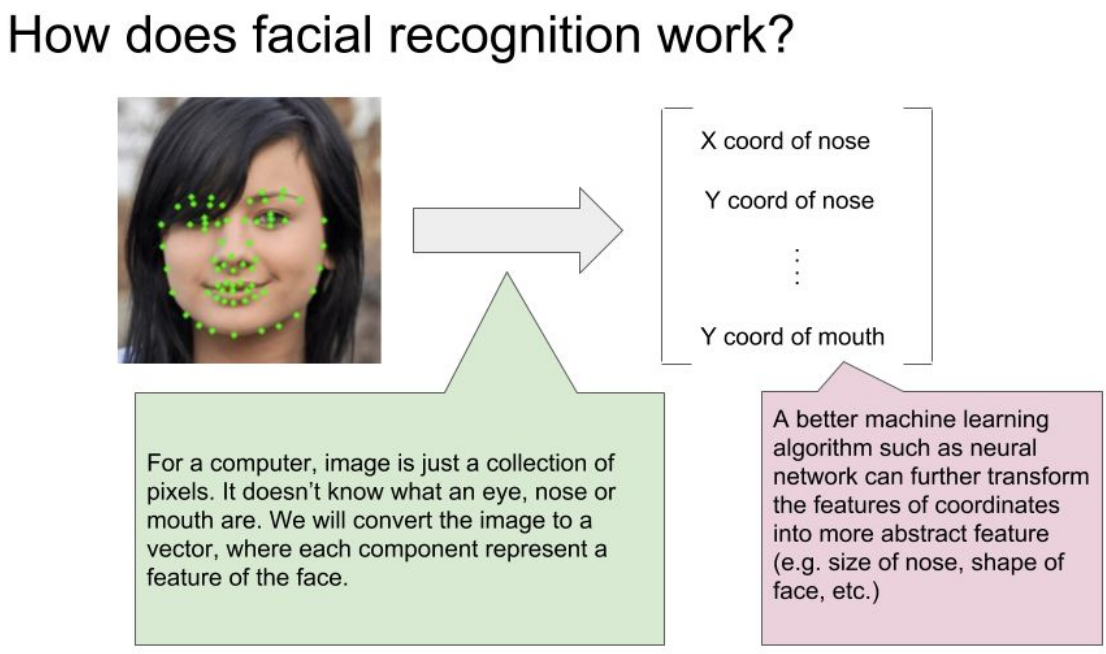
After we create the model, given a face of a person, the machine identifies if a person is a male or a female.



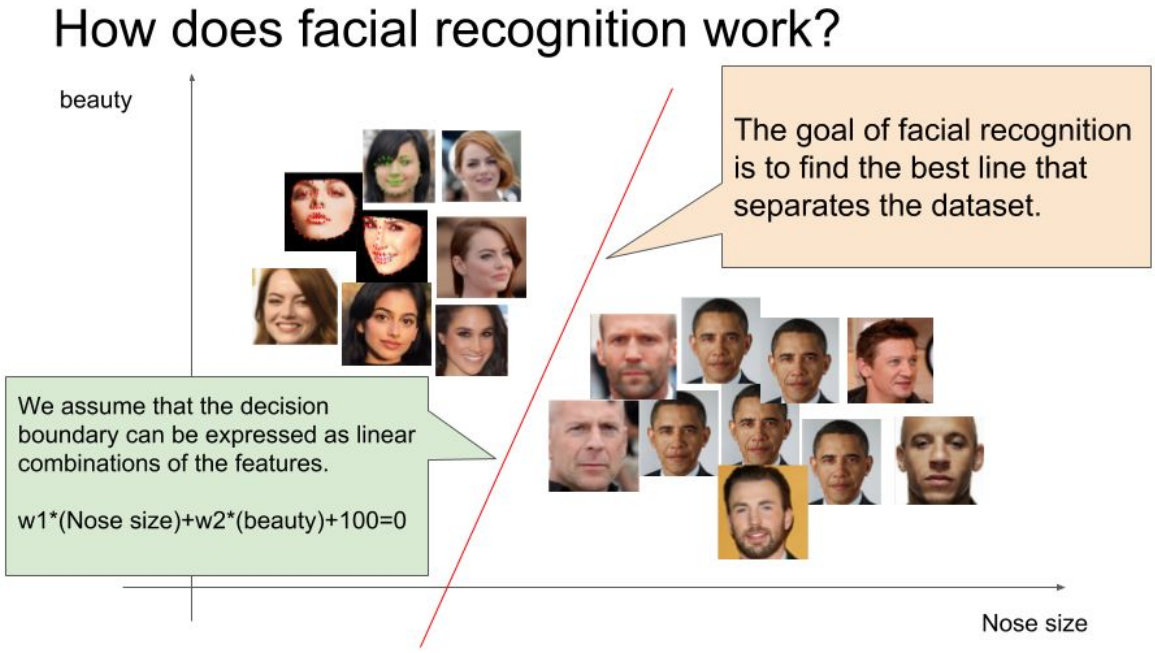
3. Classification

3.1 Feature

Although there are much better algorithms for facial recognition such as CNN, we will use logistic regression to grasp how machine learning works.



3.2 Decision Boundary



3.2 Hypothesis

The hypothesis we are making is that, linear combination of the features somehow represents if a person is male or female. But we want to know the probability of a given face being female. We can squeeze this value into (0, 1) with sigmoid function.

$$P(y = 1;\ x,\ \theta) = \sigma(\theta^T x)$$

3.3 Maximum Likelihood Estimation

We want to maximize the likelihood of observing what we observe, which can be expressed as the follow.

$$L(\theta) = P(y^{(1)},\ y^{(2)},\ \dots,\ y^{(n)};\ x,\ \theta) = P(y^{(1)};\ x,\ \theta)P(y^{(2)};\ x,\ \theta)\dots P(y^{(m)};\ x,\ \theta)$$

Notice P(y) can be expressed as the follow (since y is either 0 or 1):

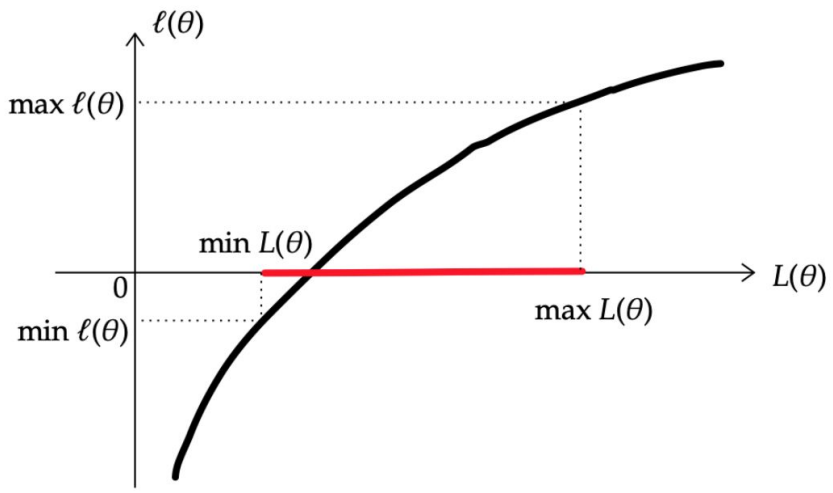
$$P(y^{(1)};\ x,\ \theta) = \sigma(\theta^T x^{(1)})^{y^{(1)}}\left(1 - \sigma(\theta^T x^{(1)})\right)^{1-y^{(1)}}$$

Therefore, the likelihood function would be

$$L(\theta) = \prod_{i=0}^{m} \sigma\left(\theta^T x^{(i)}\right)^{y^{(i)}} \left(1 - \sigma\left(\theta^T x^{(1)}\right)\right)^{1-y^{(i)}}$$

We are not going to prove it by it turns about this likelihood function is concave. To find maximum of the likelihood function, we will take the derivative of the function later. However, taking the derivative of products is nasty. So we will take the logarithm of the likelihood function. We can do this because logarithm is monotonically increasing function. Therefore, the argmax of L(θ) is same as argmax of $l(\theta)$.

$$\arg\max L(\theta) = \arg\max \ell(\theta)$$



$$\ell(\theta) = \log L(\theta) = \sum_{i=0}^{m} y^{(i)} \log\left(\theta^T x^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - \sigma\left(\theta^T x^{(1)}\right)\right)$$

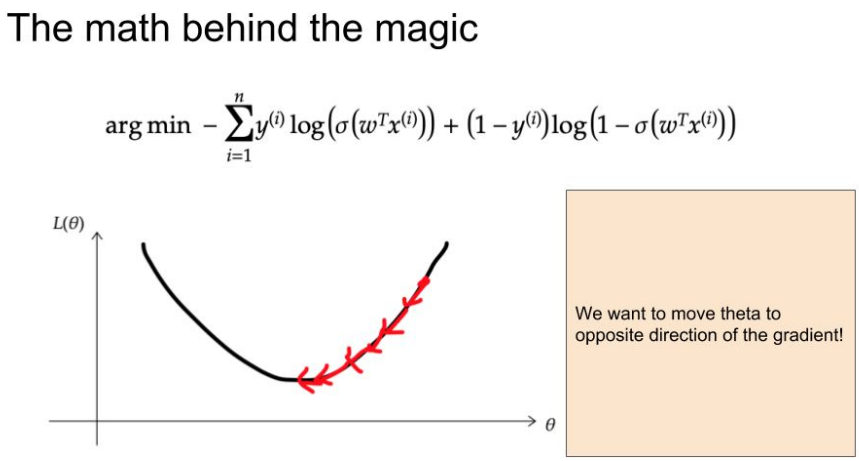
Now we want to compute the argmax of $l(\theta)$. This type of problem is called optimization problem. By convention, we generally want to minimize things in optimization. Therefore, instead of maximizing the log likelihood function, we can minimize the cost function, which is defined as the follow:

$$J(\theta) = -\ell(\theta) = -\sum_{i=0}^{m} y^{(i)} \log\left(\theta^T x^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - \sigma\left(\theta^T x^{(1)}\right)\right)$$

So the optimal value of θ will give us the best decision boundary.

3.4 Gradient Descent

We will use computational method to find the optimal value of θ as it turns out there is no closed-form solution for this. The idea of gradient descent is that we gradually modify the parameter based on the gradient of the function with respect to the parameter.



Notice we want to update the parameter in the opposite direction of the gradient.

Now we can take the derivative of the cost function, and use it for gradient descent.

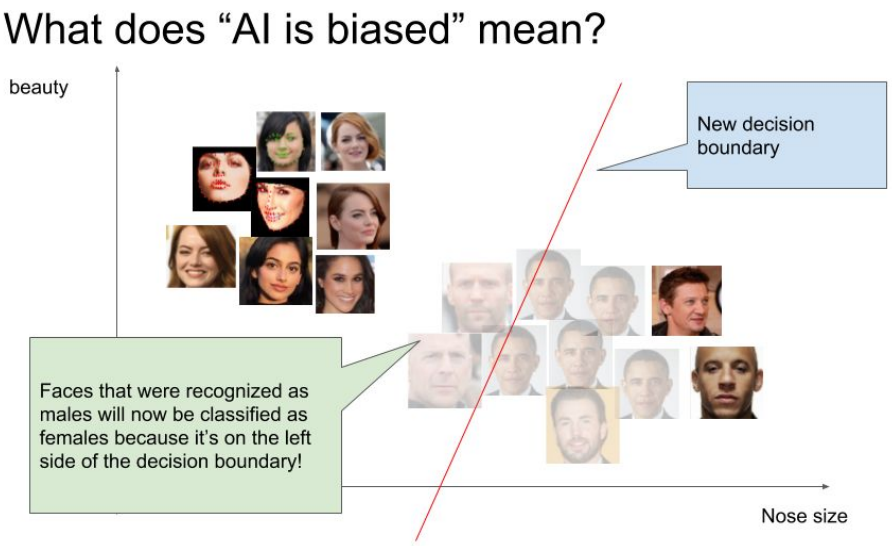
$$
\begin{aligned}
\frac{\delta J}{\delta \theta} &= -\sum_{i=1}^{n} \frac{\delta}{\delta \theta}\left(y^{(i)} \log\left(\sigma\left(\theta^T x^{(i)}\right)\right) + \left(1-y^{(i)}\right)\log\left(1-\sigma\left(\theta^T x^{(i)}\right)\right)\right) \\
&= -\sum_{i=1}^{n}\left[\frac{\delta}{\delta \theta}\left(y^{(i)} \log\left(\sigma\left(\theta^T x^{(i)}\right)\right)\right) + \frac{\delta}{\delta \theta}\left(\left(1-y^{(i)}\right)\log\left(1-\sigma\left(\theta^T x^{(i)}\right)\right)\right)\right] \\
&= -\sum_{i=1}^{n}\left[y^{(i)} \frac{\delta\sigma\left(\theta^T x^{(i)}\right)}{\delta \theta} \frac{\delta}{\delta\sigma\left(\theta^T x^{(i)}\right)} \log\left(\sigma\left(\theta^T x^{(i)}\right)\right) + \left(1-y^{(i)}\right) \frac{\delta\sigma\left(\theta^T x^{(i)}\right)}{\delta \theta} \frac{\delta}{\delta\sigma\left(\theta^T x^{(i)}\right)} \log\left(1-\sigma\left(\theta^T x^{(i)}\right)\right)\right] \left(\because \textit{chain rule}\right) \\
&= -\sum_{i=1}^{n}\left[\frac{y^{(i)}}{\sigma\left(\theta^T x^{(i)}\right)} \frac{\delta\sigma\left(\theta^T x^{(i)}\right)}{\delta \theta} - \frac{\left(1-y^{(i)}\right)}{1-\sigma\left(\theta^T x^{(i)}\right)} \frac{\delta\sigma\left(\theta^T x^{(i)}\right)}{\delta \theta}\right] \\
&= -\sum_{i=1}^{n}\left[\left(\frac{y^{(i)}}{\sigma\left(\theta^T x^{(i)}\right)} - \frac{\left(1-y^{(i)}\right)}{1-\sigma\left(\theta^T x^{(i)}\right)}\right)\frac{\delta\sigma\left(\theta^T x^{(i)}\right)}{\delta \theta}\right] \\
&= -\sum_{i=1}^{n}\left[\left(\frac{y^{(i)}}{\sigma\left(\theta^T x^{(i)}\right)} - \frac{\left(1-y^{(i)}\right)}{1-\sigma\left(\theta^T x^{(i)}\right)}\right)\frac{\delta\sigma\left(\theta^T x^{(i)}\right)}{\delta\left(\theta^T x^{(i)}\right)} \frac{\delta\left(\theta^T x^{(i)}\right)}{\delta \theta}\right]\theta \\
&= -\sum_{i=1}^{n}\left[\left(\frac{y^{(i)}}{\sigma\left(\theta^T x^{(i)}\right)} - \frac{\left(1-y^{(i)}\right)}{1-\sigma\left(\theta^T x^{(i)}\right)}\right)\sigma\left(\theta^T x^{(i)}\right)\left(1-\sigma\left(\theta^T x^{(i)}\right)\right)x^{(i)}\right] \\
&= -\sum_{i=1}^{n}\left[\left(y^{(i)}\left(1-\sigma\left(\theta^T x^{(i)}\right)\right) - \left(1-y^{(i)}\right)\sigma\left(\theta^T x^{(i)}\right)\right)x^{(i)}\right] \\
&= -\sum_{i=1}^{n}\left[\left(y^{(i)} - y^{(i)}\sigma\left(\theta^T x^{(i)}\right) - \sigma\left(\theta^T x^{(i)}\right) + y^{(i)}\sigma\left(\theta^T x^{(i)}\right)\right)x^{(i)}\right] \\
&= -\sum_{i=1}^{n}\left(y^{(i)} - \sigma\left(\theta^T x^{(i)}\right)\right)x^{(i)} \\
&= \sum_{i=1}^{n}\left(\sigma\left(\theta^T x^{(i)}\right) - y^{(i)}\right)x^{(i)}
\end{aligned}
$$

3.5 Algorithm Overview

First we will randomly pick a parameter and compute the cost function. Then, we will compute the gradient of the cost function. Next, update the parameter based on the gradient we just computed. We repeat this process until the parameter stops changing.

4. Biased Dataset

Now we roughly know how facial recognition works, let's think about the case where the training dataset is biased. Specifically, consider we have many female's face data, but not many male's face data.



The decision boundary will selected in such a way that it splits the dataset well. Notice previously correctly classified male will no longer correctly classified with this new model.

5. Solution

The solution is to make sure the dataset is not biased before training the model. Although it might seem as simple as it sounds, it it not. People are the ones who collect data, and they are more or less biased, which can affect the dataset they collect.

6. Conclusion

When people say AI is biased, what they really mean is that the dataset the model was trained with is biased.

References

1. “Amazon Face-Detection Technology Shows Gender and Racial Bias, Researchers Say.” CBS News, CBS Interactive, 26 Jan. 2019, www.cbsnews.com/news/amazon-face-detection-technology-shows-gender-racial-bias-researchers-say/.
2. Daumé, Hal. “A Course in Machine Learning.” A Course in Machine Learning, http://ciml.info/dl/v0_9/ciml-v0_9-ch06.pdf.
3. Namee, Mac B et al.“The problem of bias in training data in regression problems in medical decision support.” www.scss.tcd.ie/publications/tech-reports/reports.00/TCD-CS-2000-58.pdf.