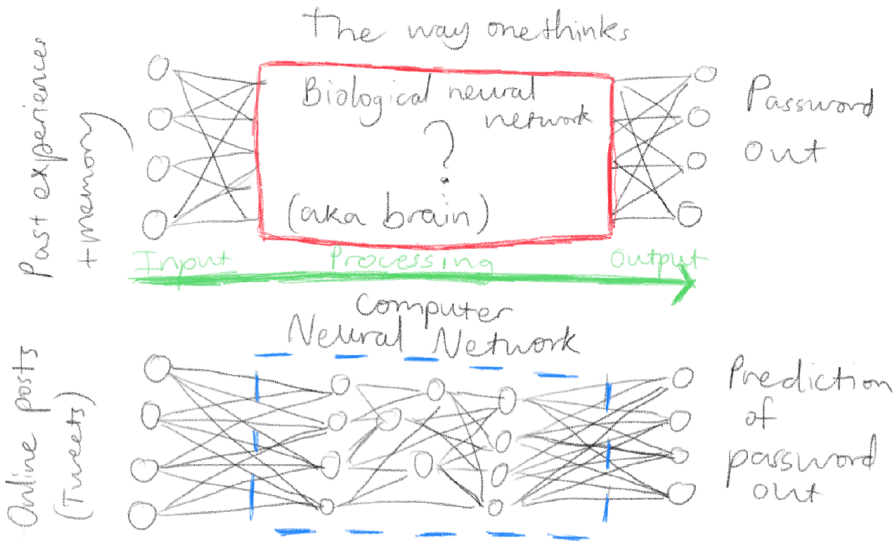
Modeling the way one thinks

PPNN - Password Prediction Neural Network

How do you define a human individual? Their body parts? Their cells and DNA? Their brain? I believe that an individual is defined by the way they think and act, by the combination of their past experiences and how they handle information. By extension, with enough data, we can model how a person consciously thinks.

My project is an early proof of concept for this idea. I chose to predict passwords:

- Because everyone thinks consciously about their password
- To highlight the importance of privacy and and danger of posting too much personal information online
- To show that passwords alone are no longer sufficient account security for most since it could be guessed and especially so if the guessing could be automated



As more data on someone is collected, the closer the input will be with their past experiences, and thus the 2 diagrams converges, allowing for more accurate models of their thinking process

In this project, I fed an LSTM neural network all of Elon Musk's tweets between 2012 and 2017, teaching it to generate passwords based on the tweets.

First prototype:

```
got mars idea is obvious (many things allow battery
dragon, our next-generation is the best
spend some press q&a on autopilot issued by is on t
launch complex 39-a. this is
solar power to cost less and work better with a the
accelerating. wish to (nor could i)
```

At this point, it was simply completing phrases a user has to manually input.

```
I uploaded a YouTube video -- Nike Sb Dunk High Pa
I uploaded a YouTube video -- New Hotrod and Cease
I uploaded a YouTube video -- Hotrod Goes To Starbu
I uploaded a YouTube video -- Nike Sb August Shipme
```

Tweets data for classification

```
rocketlandingworking
ordergivessee
getenoughzero
idiotcriticismsavoid
launchpadtesla
concerndualmotor
goodperformanceappreciation
likelotless
interestplacesworking
flightappreciationstern
```

On the left is the output from the final prototype, which had spaces and stop words removed to make it more like passwords, and had a multistep training system to increase the importance of certain tweets in a date range, which should increase accuracy when it is known when the target made the password.

One of the biggest roadblocks I ran into was the lack of data that I could legally access. I trained my network on only Elon Musk for this reason, since I could only find datasets of tweets from him and Donald Trump. I also tried to make a network to classify a user based on their tweets. It obtained an accuracy of 40%, which seemed impressive until I tried using it with real world data. It failed most of the time, since although the dataset was large, it only had fewer than 10 tweets from each user, and many of the tweets used to train the neural network were using a template (see image above right), inflating the reported accuracy.

Conclusion

Although my neural networks did not perform as well as I had hoped, I believe that it was enough as a proof of concept. A malicious actor would not have the roadblocks that I had, since they could purchase password databases on the black market and use it to make the network much more accurate. However, users can still protect themselves by using multi-factor authentication, limiting the amount of data they share and what they post about online. Knowing is winning half the battle, and the more data you share, the more you lose.

View the code on GitHub: https://github.com/o4ugDF54PlqU/PPNN