

nba-wins-nnetwork

Data

I choose to use the [sportsdataverse NBA API](#) to get NBA box scores from 2017-2022 (about 15000 box scores) to predict if the team won or not. This had multiple stats like `team_score`, `points_in_paint`, `three_point_field_goals_made`, etc. I did some cleaning and removed any non-integer/float categories (like team logo/team color/team name). Also redundant categories like `team_turnovers`, `total_turnovers`, and `turnovers` I simplified. I turned the bool target variable (`team_winner`) into an integer column instead represented by 1 (true) and 0 (false).

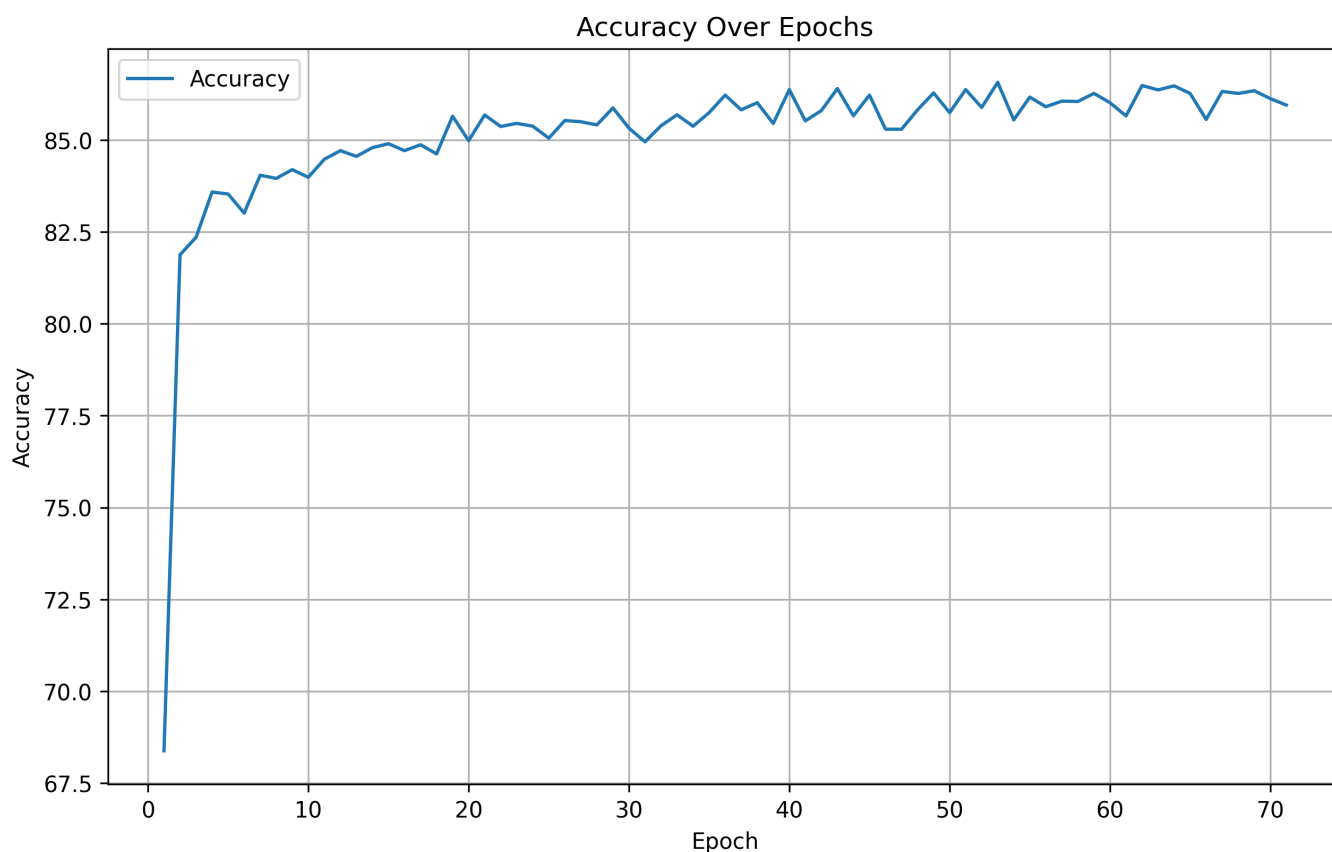
Model And Problem

I used Python's other machine learning package, `pytorch`, for this. I created a neural network that took in 23 feautres, used 3 hidden layers(linear and `relu` activation functions), and outputted whether that team won or not (the final linear layer does this by having just one output). I did this in Python as a full on program, creating 3 files (`main.py`, `model.py` where I defined my nn module, and `dataapi.py`), rather than just an exploratory problem in Colab (Only because this is a small enough problem I could run it on my regular CPU). I did end up also running it in Colab for convenience sake but the code here will run on your own computer too as long as you `pip install requirements.txt`. In Pytorch, there isn't a simple early stopping method, so I implemented if statements that measured whether the current `val_loss` was less than the `best_val_loss` so far, and reset the patience countdown variable, or subtracted 1 from that patience.

Results

Best Result so far was 86.48% accuracy against the validation set. This means it was 86.48% accurate over predicting whether a team won or not based on their box score. Here's a graph of the model's accuracy on

it's best run so far over its epochs.



Ethical Implications

There's no real ethical implications since I'm just taking team's boxscores for each game, and predicting whether they won or not (it'd be different if I was trying to predict a player from a boxscore). No personal information is involved. No data that isn't public.

Future Work

- Continued feaature refinement (which stats correlate best with wins)
- Transform the data to predict total wins for a season for a team
- Automatic scripts that grab current data to allow for pythagorean W-L predictions
- Add more data

Links to work

- [Github repo](#)
- [Colab-exploration](#)

Helpful Links

- [Neural Network with Pytorch](#)