Happy Cities

Allen T Jones
atjones@stanford.edu

Problem:

As humans, most everything we do is centered directly or indirectly around the pursuit of happiness. Even when we do not think in those terms, it is the primary driving force in our lives once basic survival needs are met. When I picked the city I would move to after getting the offer for my current job, I did not ask myself “which city will make me the happiest”. I looked at the commute time, the ratio of house to dollar I could afford, the crime levels, the green spaces, air quality, and more. In the end, every one of those points to happiness as it’s end goal. What if you could just look an introduction to a city and see how happy it was? What if you could use that information to gain hints at how to improve happiness in a city you were in?

Oracle:

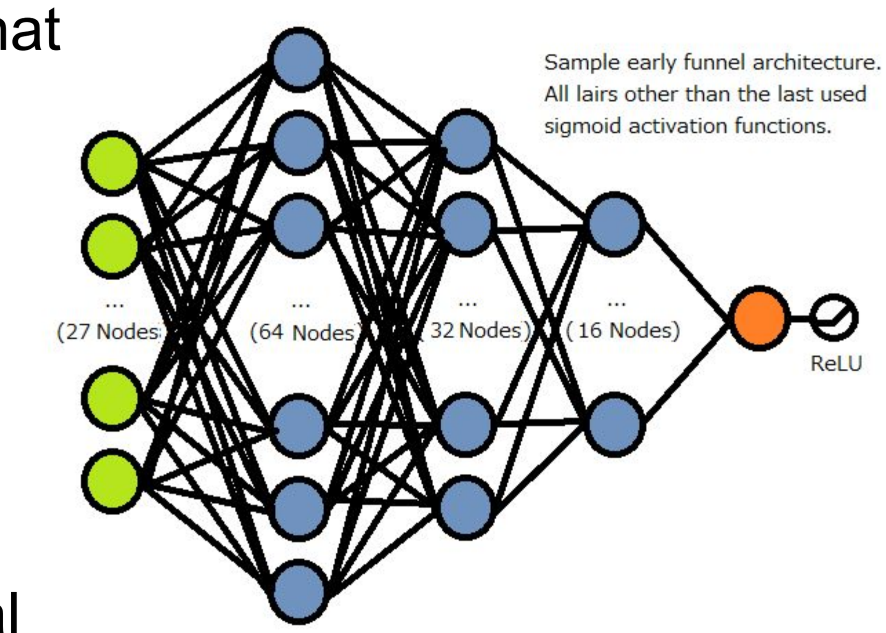
Earlier this year, website WalletHub did an analysis ranking the “Happiest Cities in America” using 31 indicators and 182 of the largest cities in America. They used data from a wide variety of sources, many of which are not available for smaller cities. This covered a broad selection of things such as depression levels, sports participation, divorce rates, and whether people get enough sleep.

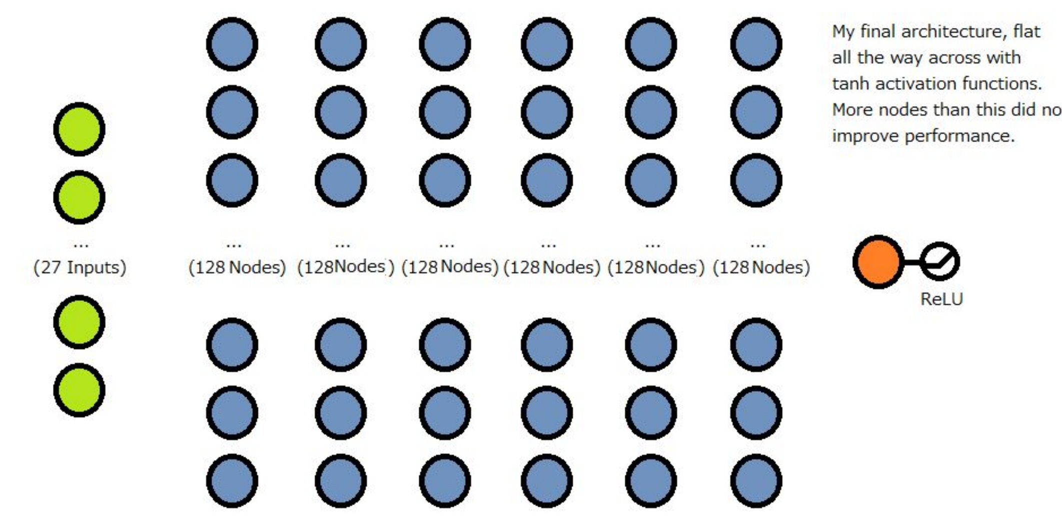
Proposal:

Using the scores assigned by the oracle analysis, I hoped to map data that was easily available from a centralized source to happiness scores with a supervised learning model. In this case, census data is the most easily available large database, with a variety of factors that seem likely to correlate to happiness such as commute time and income levels. In total there were 27 features which I extracted from the census data and normalized.

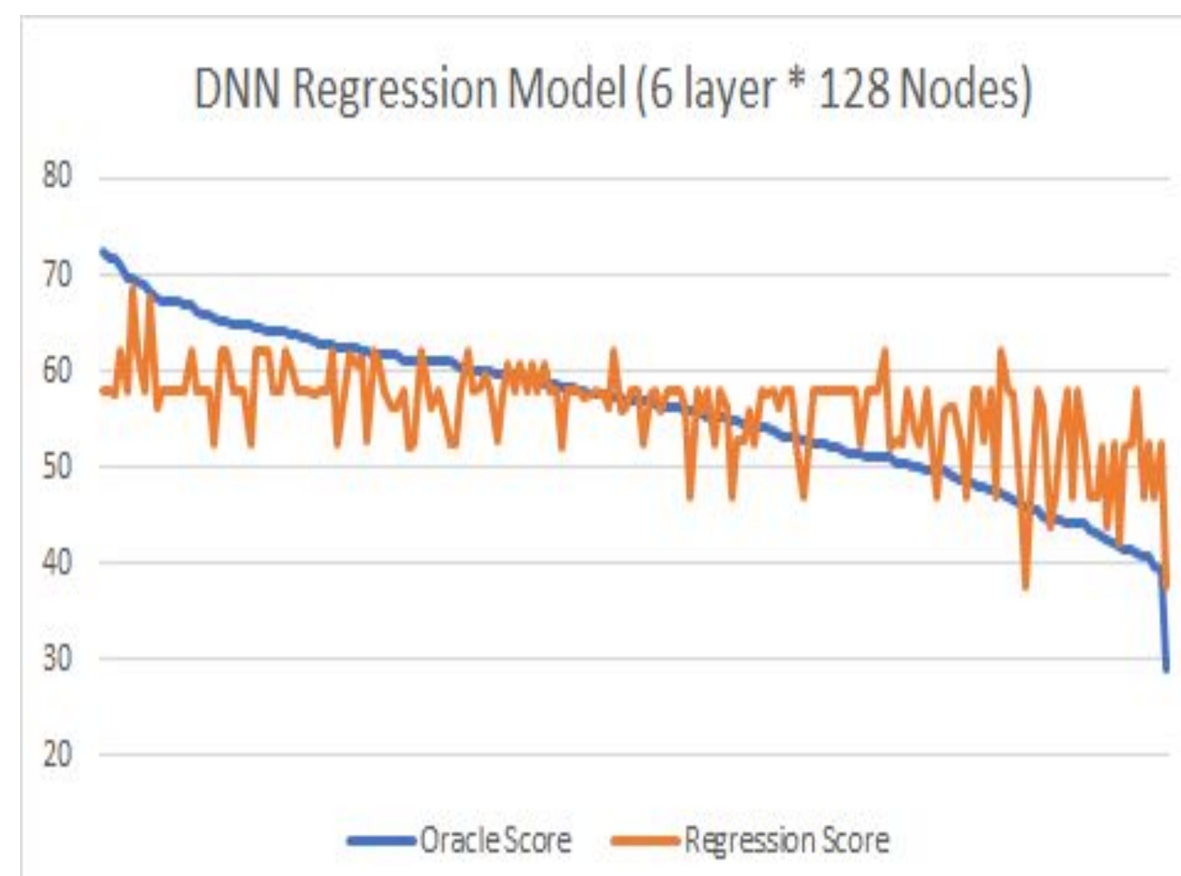
Implementations:

In pursuit of a model that actually worked I tried a variety of shapes, depths, and activation functions. Given that I only wanted one output, my initial inclination was to try a funnel shaped deep neural network, but I found that flat networks worked better for every case I tested. Layer depths of more than 128 nodes showed no improvement, and tanh activations for the internal layers outperformed other activation functions by a small margin.

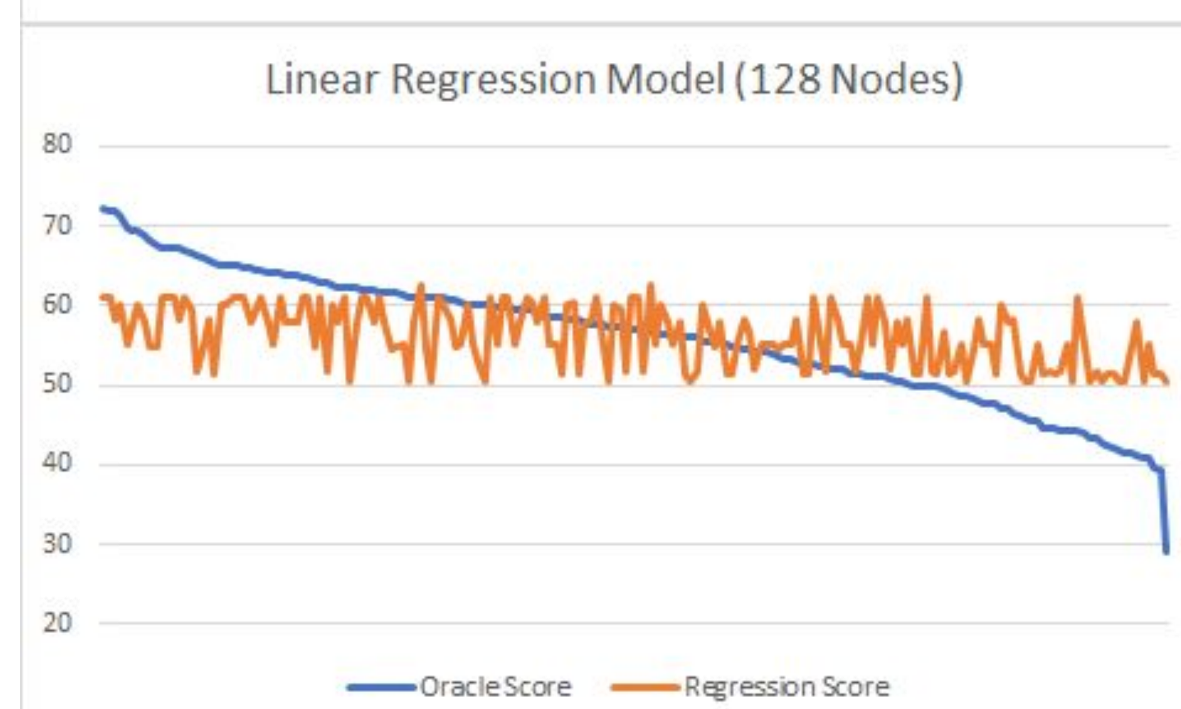




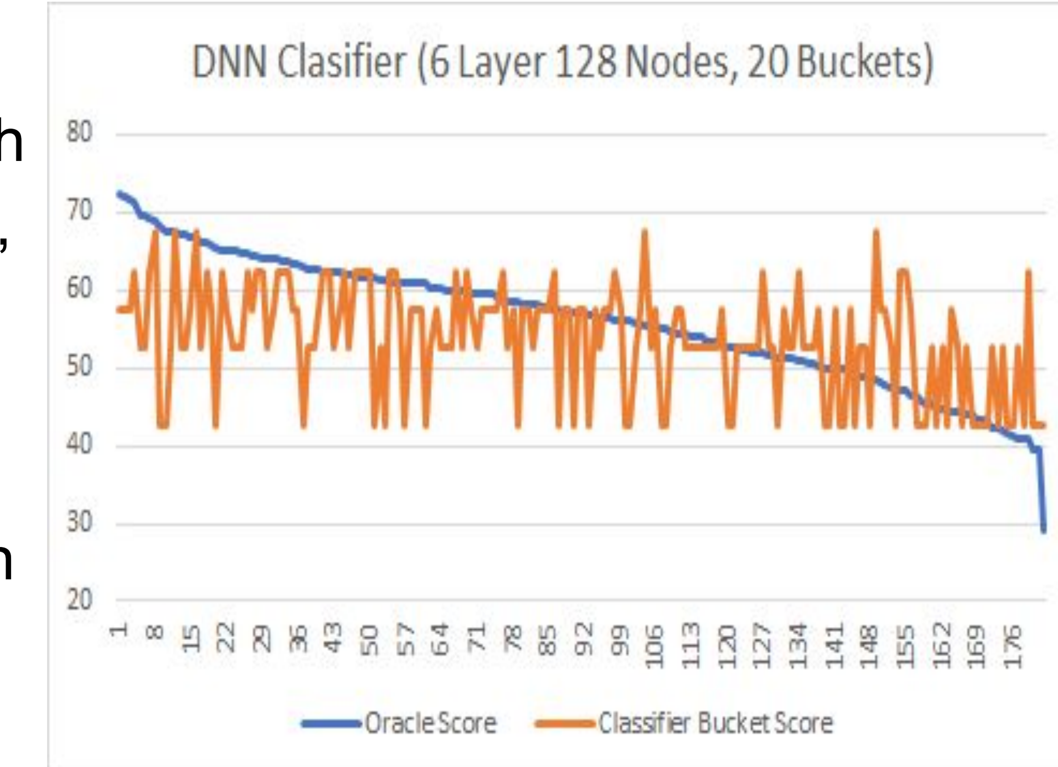
My best performing model was woefully inaccurate, showing at best a rough correlation between the census data and the oracle ranking.



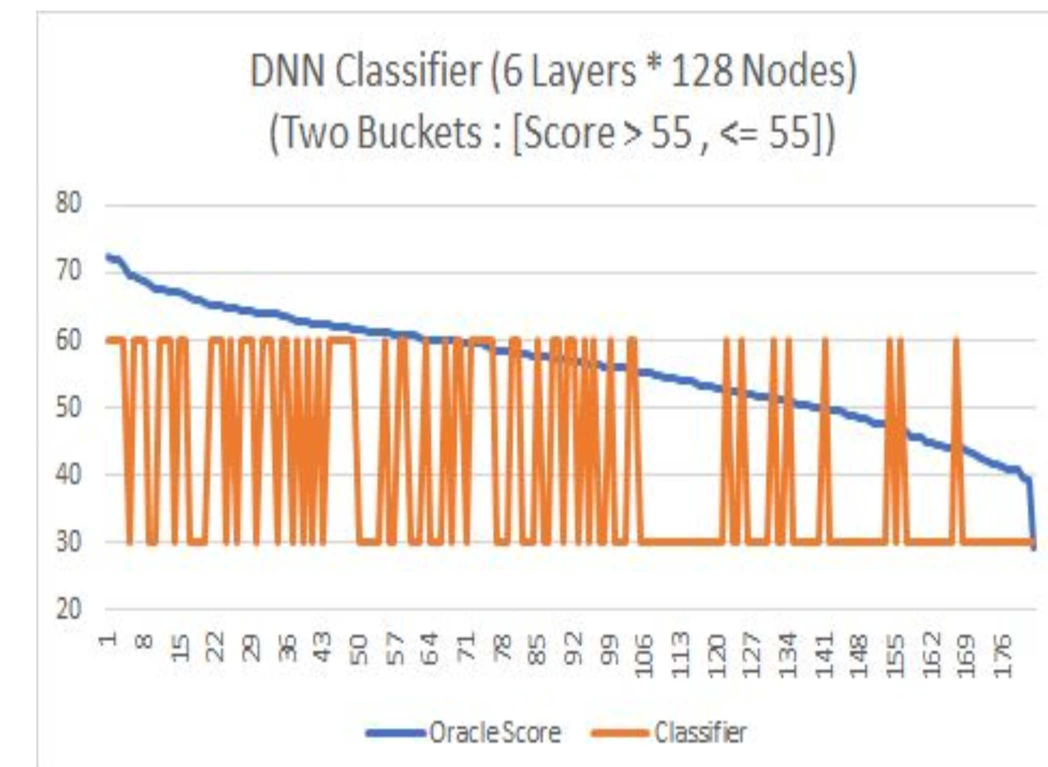
I also tested a single layer linear model to compare to the neural network performance.



In an attempt to find an alternate approach to analyzing the data, I switched from regression to classification, using bins to group score in 5 points increments.



Hoping to extract an additional piece of information from a simpler problem that might be used to as an additional input to the main network, I attempted to see if I could at least train a network to recognize whether the query was in the top or bottom half of results. This again proved difficult, with a successful classification rate of approximately 70%.



Results:

Happiness is ineffable. It’s hard to classify or quantify. More than anything, it’s pursuit is not so simple as to allow us to plug in census data and pop out a best place to live.

There is no correlation between basic census data and happiness as quantified by my oracle source. At this point I believe it is impossible to improve my results without more data points to provide as inputs in addition to the census data.

References:

https://wallethub.com/edu/happiest-places-to-live/32619/
https://www.pnas.org/content/107/38/16489#T1