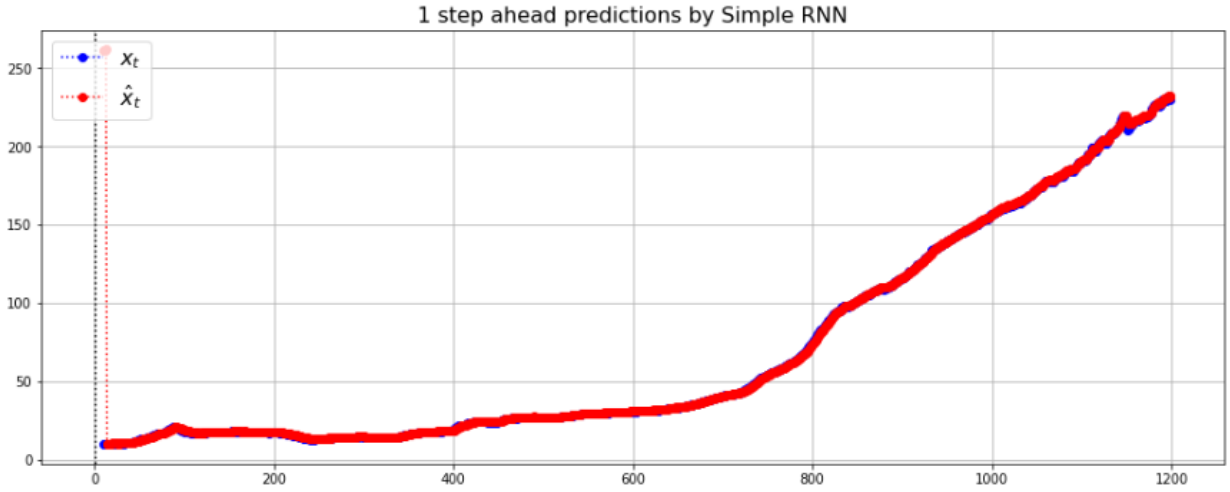
Factors to predict CPI using neural network

Mofei Wang

Due to Covid19, we had a huge inflation and the consumer price index raised by almost 12% for the past year. Basically, Consumer Price Index is an index to measure the average change over time in the price paid by consumers in the market. (For both good and services and this research will mainly focus on the US CPI). The inflation impact our everyday life; therefore, I am interested in how can I use the methods learned in our class to try to predict the consumer price index.

CPI(Monthly from 1913 to 2020)

Here, I used CPI itself to predict the future CPI. The one step ahead prediction simple RNN is implemented and here is the result :

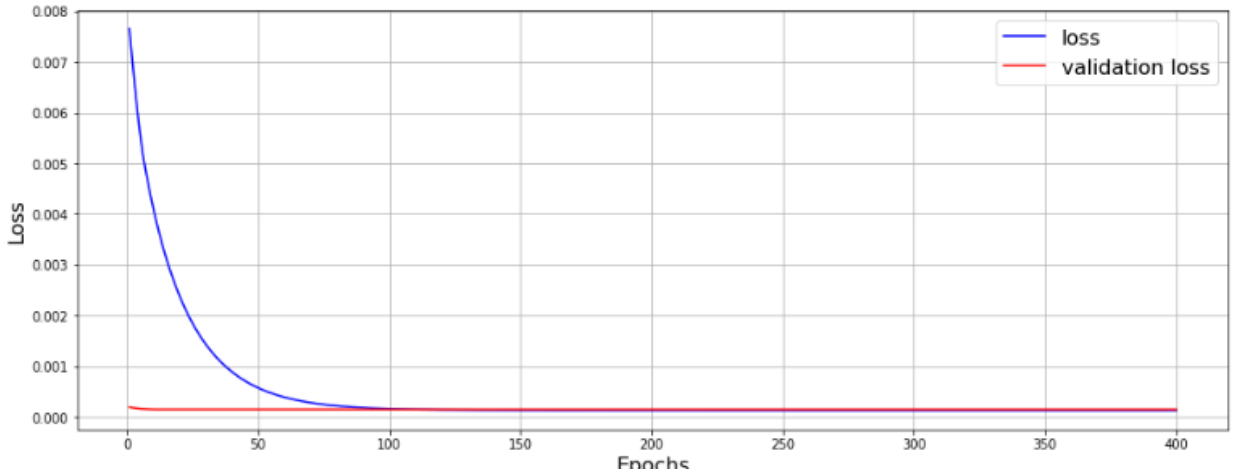


We could see that the CPI grew from around 10 to nearly 240 today in 108 years. This result above seems to have a high loss. One reason could be I do not have enough data. Another reason could be, there are a lot of factors that might influence CPI. For example, wars, economic crisis, or Covid (that is why I did not put data after year of 2020 in) could greatly impact CPI.

Population/RealGDP(Yearly from 1929 to 2020)

Then I tried to put some factors in. First I tried to use Population and RealGDP(not nominal gdp, adjusted to 2012) as factors to predict CPI by a simple neural network with 4 hidden layers. However, for those factors, I could only found data yearly and the data is really limited which makes the prediction hard to do, here is the model summary and the loss/validation loss graph:

```
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_3 (Dense)             (None, 32)                96

 dense_4 (Dense)             (None, 32)                1056

 dense_5 (Dense)             (None, 64)                2112

 dense_6 (Dense)             (None, 32)                2080

 dense_7 (Dense)             (None, 1)                 33

=================================================================
Total params: 5,377
Trainable params: 5,377
Non-trainable params: 0
_________________________________________________________________
```

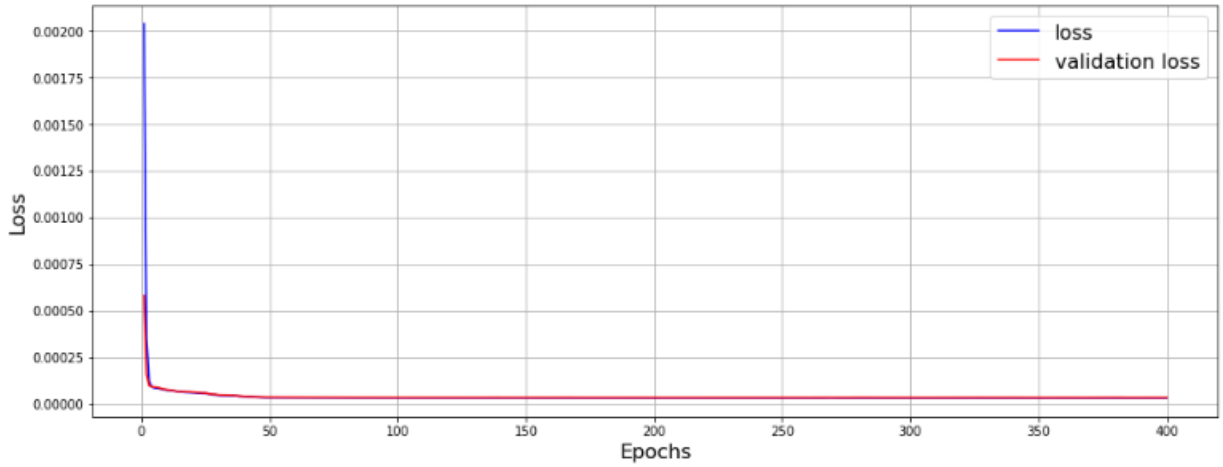


Note the output is scaled to 1/10000. The minimum validation loss here is **0.00015191052807494998**.

Gold price(Monthly from 1913 to 2020)

Secondly, I tried to use Gold price as a factor to predict CPI. The reason is that it is easy to find the monthly data or even daily data for Gold price and if there is a large inflation gold price usually increase. If Gold price predicts CPI well, we could get larger data sets to train a better model. (we have been using CPI for less than 200 years, but there are evidence we have been using Gold for nearly 6000 years) On the other hand, we could use Gold price to "predict" CPI backward to help researching the commercial life of our ancients. Here are the model summary and loss graph:

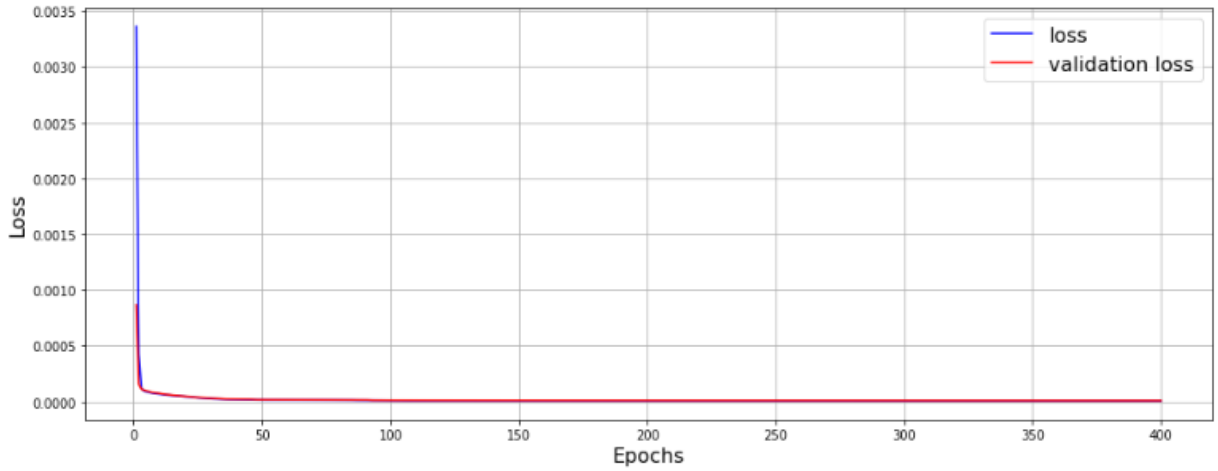
```
Model: "sequential_6"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_23 (Dense)            (None, 32)                64

 dense_24 (Dense)            (None, 32)                1056

 dense_25 (Dense)            (None, 64)                2112

 dense_26 (Dense)            (None, 32)                2080

=================================================================
Total params: 5,312
Trainable params: 5,312
Non-trainable params: 0
_________________________________________________________________
```



Note the output is scaled to 1/10000. The minimum validation loss here is **3.2560117688262835e-05.**

CPI/Gold price(Monthly from 1913 to 2020)

At last, I tried to use CPI and Gold price all together to predict CPI to compare with the one only uses Gold price as the factor. Here is the loss graph:



Note the output is scaled to 1/10000. The minimum validation loss here is **9.2999471235089e-06.** This choice of factor actually has the lowest validation loss. But again, CPI has a limited database, so I would prefer to gather all the gold price data to train the model to predict CPI.

Youtube presentation link

https://youtu.be/CuhhhEz9JN0

Configuration and data source

Since I have a relatively small dataset here, I used the online notebook deepnote to run my codes.(It should take less than 5 minutes to run in local Jupyter notebook) For the data below, some of the data are combined from several websites as most of them only have part of the data I needed.

CPI:
https://datahub.io/core/cpi-us#resource-cpi-us_zip
https://www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from-1913-to-2008/
https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1913-

Population:
https://www2.census.gov/programs-surveys/popest/tables/1900-1980/national/totals/popclockest.txt
https://www.macrotrends.net/countries/USA/united-states/population

GDP:
https://www.thebalance.com/us-gdp-by-year-3305543

Gold prices:
https://datahub.io/AcckiyGerman/gold-prices
https://www.indexmundi.com/commodities/?commodity=gold&months=60