Final Project Demo

Allen Wang

2022-04-10

My initial goal of this project is to predict the unemployment rate of the Los Angeles County, so I need to build up my own data set. Currently, my data set is over ten predictor variables and the unemployment rate from 1990 to now. I will keep updating my data set to meet a minimum of 25 potential predictor variables. All variables are economic indexes presented monthly, giving me about 380 observations. These variables are all extracted from FRED and US BUREAU OF LABOR STATISTICS

Basically, the predictor variables can be separated into two parts: local LA economics indicators and National economics indicators, which are all numerical variables. The selecting of predictor variables requires specific economics analysis, so I only gather a few potential predictor variables for now.

Yes, there exists missing data for some predictor variables from 1990 to now. For some reason, one of the variables *(California Leading Index)* was not updated anymore in 2019. I may exclude those variables without enough time range, but current data seems to be fine. Luckily, there is no missing value in the middle of time range, and this offers me to train model in a long period of time starting from 1990.

An overview of my research questions

As I have mentioned, I am interested in predicting the unemployment rate of Los Angeles. Here is the link of my PROPOSAL. As the response variable, the unemployment rate is *people above a specified age (usually 15)[2] not being in paid employment or self-employment but currently available for work during the reference period*(wiki). My model will be predictive, and it will be achieved by a regression approach.

Variables especially related to unemployment should be very useful. For example, the number of employee of *Trade, Transportation, and Utilities* section can be used to forecast the cyclical change of the unemployment rate. Also, general economics index, such as LA CPI and national GDP, can be very helpful.

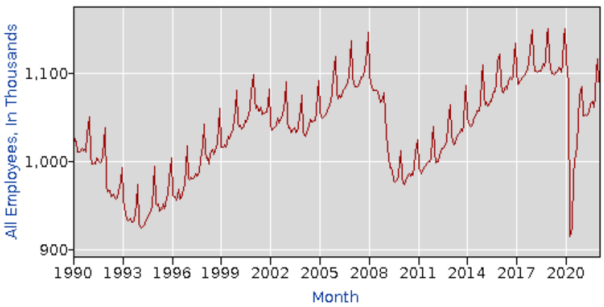

Figure 1: a local picture

Here is a brief overview of the data set.

```r
econdata = read.csv("CALOSA7URN.csv")
head(econdata)
```

```
##       DATE unemployment_rate home_price_index Labor_force CPI.U    M1
## 1 1990/1/1               5.9         100.4712     5864160 132.1 795.4
## 2 1990/2/1               5.6         100.7616     5891997 133.6 798.1
## 3 1990/3/1               5.4         100.9928     5902033 134.5 801.5
## 4 1990/4/1               5.5         100.9904     5863163 134.2 806.1
## 5 1990/5/1               5.4         100.3946     5879518 134.6 804.2
## 6 1990/6/1               5.4         100.1110     5885104 135.0 808.8
##   per_con_exp fed_fund_eff_rate inte_rate curr_incircu Cali_lead_index
## 1      3730.7              8.23         7      257.126            1.87
## 2      3728.2              8.24         7      254.662            1.30
## 3      3754.9              8.28         7      256.690            0.89
## 4      3770.0              8.26         7      259.829            0.51
## 5      3775.8              8.18         7      261.938            0.08
## 6      3804.5              8.29         7      265.640           -0.21
```

Project Timeline

(1) Data Gathering: hopefully, I can gather a satisfactory amount of data and load it at the end of the third week.

(2) Exploratory Data Analysis: I have considered the work load of this section, but it definitely will takes lots of time for learning economics indexes and reading existing papers about unemployment model and forecast. Thus, I will plan three weeks for this steps.

(3) As the course proceeds, I will have a general idea of machine learning models, and I will conduct model selection and compare during the remaining quarter.

Questions and Concerns

Consider the previous papers, the neural network is a feasible and effective model to predict the unemployment. In detail, I should utilize data from previous three months to predict the unemployment next month. This method may require extra knowledge about working with time-series analysis. So, I wonder if this course will cover how to analyze time-series data.

Will the instructional team provide additional help before the deadline of the final project?