Learning With Kernel Machines   
Map Your Own Problem

Instructions:

In this part of the course project, you will map your own problem taking the ERM approach. You will specifically describe the particulars of the problem you would like to solve. Answer the following questions within the provided textboxes.

What is the problem space/context?

|  |
| --- |
| Cryptocurrencies have become very popular in recent years. They can be used to pay for services and also as speculative investments. More investors and traders start considering including cryptocurrencies in their own portfolios or investment strategies. A critical problem for many crypto investors and traders is how to predict the price or return of a cryptocurrency. |

What do you want to learn?

|  |
| --- |
| I would like to learn about predicting Bitcoin’s next-day return in USD with given features such as US interest rates, US inflation, Bitcoin’s previous day/month’s returns, Bitcoin’s daily trading volume, and NASDAQ-100 composite return. |

What are the constraints/limitations of your data?

|  |
| --- |
| Currently, I can think of the following major constraints for predicting Bitcoin’s return:   1. **Small sample size.**The earliest Bitcoin price data on most online sources are after 2009. Therefore, we have less than 5500 sample sizes of Bitcoin daily return, and we might need to split the data into training and testing sets to validate the results. 2. **Different data frequency.** Some factors/features may not have the same frequency as Bitcoin daily return. For example, US inflation data published by US Bureau of Labor Statistics is updated on a monthly basis. We might need to transform the monthly data into daily basis or impute the daily data based on monthly data. 3. **Change of market scenario.** The underlying association of Bitcoin with the features may change as the market fundamentally changed. We might need to apply change point detection or some intuition to determine the relevant data for model training. |

What algorithm, loss function, and regularizer will you use? Why?

|  |
| --- |
| For loss function, since Bitcoin daily return is continuous and not noisy, squared loss can be used.  For regularizer, both L1 and L2 can be used, and I prefer to use L2 regularization since I would like to reduce model variance and a sparse solution is not required by the problem.  Combining squared loss and L2 regularization together, I would select **Ridge regression** for predicting Bitcoin’s daily return price. |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*To submit this assignment, please refer to the instructions in the course.*