

Contents

1 Objective	3
2 What is Hidden Markov Model?	4
3 Implementation details	6
3.1 Programing Language	6
3.2 Packages	6
3.3 Strategy	7
3.4 Data	7
4 Backtesting Code	9
4.1 FinancialData	9
4.2 BacktestBase	10
4.3 Portfolio	11

5 Training model for HMM	12
6 Strategy code without using HMM	15
6.1 MovingAverageStrategy	16
7 Strategy code with using HMM	20
8 Findings	26
9 Future Work	29
10 Reference	30

Chapter 1

Objective

There is always challenge for quantitative trader to find out the frequent behaviour of financial market due to change in government policy,negative news item,regulatory environment and other macroeconomics effects. Such periods are known as Market Regime. These various regimes lead to adjustments of asset returns via shifts in their means, variances, autocorrelation and covariances. This impacts the effectiveness of time series methods that rely on stationarity. There is a clear need to effectively detect these regimes. This aids optimal deployment of quantitative trading strategies and tuning the parameters within them.

This project is an attempt to find out such market regime and accordingly adjust the strategy. The principal method used to detect market regime is known as Hidden Markov Model which is a statistical time series technique.

Chapter 2

What is Hidden Markov Model?

Before knowing about Hidden Markov Model, it is important to understand Markov Model. The Markov Model is a stochastic state space model involving random transitions between states where the probability of the jump is only dependent upon the current state, rather than any of the previous states. The model is said to possess the Markov Property and is thus **memoryless**.

Markov Models can be categorised into four broad classes depending upon the autonomy of the system and whether all or part of the information about the system can be observed at each state.

	Fully Observable	Partially Observable
Autonomous	Markov Chain	Hidden Markov Model
Controlled	Markov Decision Process	Partially Observable Markov Decision Process

If a model is both autonomous and fully observable. It cannot be modified by actions of an agent as in the controlled processes and all information is available from the model at any point in time.

If the model is fully autonomous but only partially observable then it is known as a Hidden Markov Model. In such a model there are underlying latent states and probability transitions between them but

they are not directly observable. Instead these latent states influence the observations.

The HMM is more familiar in the speech recognition community and communication systems, but during the last years gained acceptance in finance as well as economics and management science.

Chapter 3

Implementation details

3.1 Programing Language

The project has been implemented in Python version 2.7. Python being a open source language is most popular for data analysis as it has wide range of available packages for data analysis,machine learning and statistical analysis. It is very popular language and easy to use.

3.2 Packages

The list of main packages used in this projects are as follows.

1. Python (v2.7)
2. Pandas (0.18.1)
3. Numpy (1.13.1)

4. Matplotlib (2.0.2)

5. hmlearn (0.2.0)

3.3 Strategy

As the main focus is on detection of market regime ,so I have used moving average strategy. As in case of HMM we can only have partially observable data, it become very important choose your observable data very carefully. For Moving average strategy I have choosen the daily returns as observable variable. We can also have daily std deviation ,daily volatility as observable variables.

In future work, I am going to add few more strategy with different observable variable.

3.4 Data

I have taken 120 days interaday data from Zerodha for the following script.

1. NIFTY50 Index

2. HDFCBANK

3. ICICIBANK

4. KOTAKBANK

5. ONGC

6. INFY

7. RELIANCE

8. HDFC

9. LT
10. IOC
11. SBIN
12. HINDUNILVR
13. MARUTI
14. ITC
15. TCS

It is always advisable to go for interaday data to get better result for detecting market regime using HMM.

Chapter 4

Backtesting Code

All the backtesting code is written in **backtest.py** file. The file is divided into two class namely **FinancialData** and **BacktestBase**.

4.1 FinancialData

This class get all the data required for backtesting. The class has following function.

- **init** : It initialize with ticker name. Also profolio class object is initialized here.
- **get_data** : All data reading happen in this function. Here we are loading the data in the pandas dataframe from csv file for the given ticker. You can load it from database as well.
- **plot_data** : This function plot the data on given column. By default it plot on close data.

4.2 BacktestBase

This class inherit all the property of FinacialData. Which means I can access the `get_data` and `plot` function from this class as well. This class has follwing function.

- **init** : It initialize with initial amount invested, amount after all backtesting,no of trades ,units ,position and transaction cost based on broker. I am calculating brokerage charge as per zerodha brokerage. So I am not using `ftc` and `ptc`.
- **get_trade_price** : This function gives the current trade price based on number of units.Also it add the transaction cost to it.
- **get_date_price** : This functon give the price of security at current data index level.
- **get_txn_cost** : Based on the amount invested it calculates the transaction cost.This function may change based on different broker.
- **get_trade_units** : This function gives the number of units you can buy based on the available amount.
- **place_buy_order** : This function place the buy order and accordingly modify the trade details,units and available amount.
- **place_sell_order** : This function place the sell order and accordingly modify the trade details,units and available amount.
- **trade_stats** : This function give the trade status once the backtesting is completed. Here I am calculating,initial amount invested, final balance, performance,number of trades and sharpe ratio. You can calculate your own stat here.

4.3 Portfolio

This class is written in **portfolio.py**. It actually being called from **trad_stat** function for each and every symbol. This class is responsible for maintaining the portfolio details of every symbol on which we want to run the backtesting strategy. Here we can find out which symbol is performing better for our backtesting strategy and we can focus on those symbols in live trading.

Portfolio class has following methods.

- **init** : It initialize portfolio details which hold column name for our portfolio stats.
- **add_portfolio_details** : This function add the portfolio details for every individual ticker.
- **show_portfolio_details** : This function actually store the portfolio details object in the pickle file and save on disk which can be load at anytime for comparison , presentation etc.

Chapter 5

Training model for HMM

It is important to split the data into training and test data into proper ratio. As I have only 120 days intraday data it was difficult for me to have 70:30 ratio so I have decided to split it into 50:50 ratio. Please note that more the amount of data better would be the result. If you have large data the I would suggest you to go with 70:30 split which can be optimized.

The file used for training model is **hmm_regime_training_model.py**. This code actually split the data into training and test data in 50:50 ratio. To create the model **hmmlearn** package of python has been used. Please note that while creating the model it is important to choose the observable variable very carefully as on this variable model is going to be fit. Here in moving average strategy the observable variable is **daily returns**. We can also use **standard deviation, volatility, forward volatility(for options) etc.**

Once the model is created, it is dumped to a file using same **pickle** module of python. It creates file for each and every symbol and store it into **Model** directory.

The code snippet for training HMM model is as follow.

```
1 import warnings
2 import pandas as pd
3 import datetime
4 import numpy as np
5 import pickle
6 from hmmlearn.hmm import GaussianHMM
7
8 data_location = "../Data/IntraDay/"
9 train_test_split_ratio = 0.5
10
11
12 # Hides deprecation warnings for sklearn
13 warnings.filterwarnings("ignore")
14 tickerList = ["HDFCBANK", "ICICIBANK", "KOTAKBANK", "ONGC", "INFY", "RELIANCE", "HDFC", "LT", "IOC",
15              "SBIN", "HINDUNILVR", "MARUTI", "ITC", "TCS"]
16
17 for ticker in tickerList:
18     model_path = "../Models/hmm_model_"+ticker+".pkl"
19     data = pd.read_csv(data_location+ticker+"-EQ.csv", index_col='Date', names=['Date', 'Open', 'High', 'Low', 'Close', 'Volume'], skiprows=1)
20     data["Returns"] = data["Close"].pct_change()
21     data.dropna(inplace=True)
22     #Reverse data to get proper ascending order
23     data = data.iloc[::-1]
24
25     ##Split data into training and test data
26     training_data_len = int(len(data)*train_test_split_ratio) ##TO get training data
27     end_date = data.index[training_data_len]
28     start_date = data.index[training_data_len+1]
29     training_df = data[:end_date]
30     test_df = data[start_date:]
```

```

30 # print training_df.tail()
31 # print test_df.head()
32
33 ##Observable variable on which hmm model will fits
34 returns = np.column_stack([training_df["Returns"]])
35
36 # Create the Gaussian Hidden markov Model and fit it
37 # to the returns cloumns fo training data, outputting a score
38 hmm_model = GaussianHMM(n_components=2, covariance_type="full", n_iter=1000).fit(returns)
39
40 #Dump model in a file to use it later on
41 pickle.dump(hmm_model, open(model_path, "wb"))

```

Chapter 6

Strategy code without using HMM

The file used to create moving average strategy is **moving_average.py**. If we run this file , it will run the moving average strategy on all the listed security given in the list. Also, you can declare initial amount to be invested. The portfolio object will give you the final status and also save it on disk for future use.

```
1 tickerList = ["HDFCBANK", "ICICIBANK", "KOTAKBANK", "ONGC", "INFY", "RELIANCE", "HDFC", "LT", "IOC",  
               "SBIN", "HINDUNILVR",  
2             "MARUTI", "ITC", "TCS"]  
3  
4 initial_investment_amount = 10000  
5  
6 for ticker in tickerList:  
7     sma = MovingAverageStrategy(ticker, initial_investment_amount)  
8     sma.run(50, 250)  
9  
10 portOBJ = Portfolio()  
11 portOBJ.show_portfolio_details("port_without_hmm.pkl")
```

6.1 MovingAverageStrategy

This class implement the moving average strategy. The class inherits the BacktestBase which inherits FinancialData class. So, basically it has access to data as well as backtesting class.

It only has **run** method which is called to execute strategy on a given symbol and data. It copy the data into another data frame and do not touch origianl data frame.It split the data into 50:50 ratio same as we did it in training mode section. Please note that, here we only have to run the strategy on testing data only even though we are using moving average without HMM. Because if we run the strategy on whole data then it would be useless to compare the result between startegy with HMM and strategy without HMM as strategy with HMM will run on test data only as training data will be used for training model.

The complete code for moving average is as follow.

```
1 import pandas as pd
2 import numpy as np
3
4 from backtest import BacktestBase
5 from portfolio import Portfolio
6
7 train_test_split_ratio = 0.5
8
9
10 class MovingAverageStrategy(BacktestBase):
11     def run(self,SMA1,SMA2):
12         msg = 'Running SMA strategy for %s |SMA1=%d |SMA2=%d |ftc=%f|ptc=%f'
13         msg = msg%(self.ticker,SMA1,SMA2,self.ftc,self.ptc)
14         self.position = 0
15         self.amount = self.initial_amount
16         self.trades = 0
17
18
```



```

19     #Data Preparation
20     self.data_run = self.data.copy()
21     self.data_run['SMA1'] = self.data_run['Close'].rolling(SMA1).mean()
22     self.data_run['SMA2'] = self.data_run['Close'].rolling(SMA2).mean()
23     self.data_run["Returns"] = self.data_run["Close"].pct_change()
24     self.data_run.dropna(inplace=True)
25
26
27     ##Split data into training and test data
28     training_data_len = int(len(self.data_run)*train_test_split_ratio) ##TO get training
29     data
30     end_date = self.data_run.index[training_data_len]
31     start_date = self.data_run.index[training_data_len+1]
32     training_df = self.data_run[:end_date]
33     test_df = self.data_run[start_date:]
34
35     #Running for test data only
36     self.data_run = test_df
37
38     # print self.data_run.head()
39
40     #####Signals
41     Signals = pd.DataFrame(index=self.data_run.index)
42     Signals["PnL"] = 0
43     Signals["Trade"] = 0
44     Signals["Units"] = 0
45
46     for bar in range(0,len(self.data_run)):
47         if self.position == 0:
48             if self.data_run['SMA1'].ix[bar] > self.data_run['SMA2'].ix[bar]:
49                 self.place_buy_order(bar,amount=self.amount)

```

```

50         self.position = 1 #Take Position
51         Signals["Trade"].ix[bar] = 1
52         Signals["Units"].ix[bar] = int(self.units)
53     else:
54         Signals["Trade"].ix[bar] = 0
55
56     elif self.position == 1:
57         if self.data_run['SMA1'].ix[bar] < self.data_run['SMA2'].ix[bar]:
58             Signals["Units"].ix[bar] = int(self.units)
59             self.place_sell_order(bar, units=self.units)
60             self.position = 0 #Market nuetral
61             Signals["Trade"].ix[bar] = -1
62         else:
63             Signals["Trade"].ix[bar] = 0
64
65
66     ##Squaring off if holds any stock without checking any condition
67     if self.position == 1:
68         Signals["Units"].ix[bar] = int(self.units)
69         self.place_sell_order(bar, units=self.units)
70         self.position = 0 #Market nuetral
71         Signals["Trade"].ix[bar] = -1
72
73
74     ##Concatenating both dataframe
75     frames = [self.data_run, Signals]
76     self.final_dataframe = pd.concat(frames, axis=1, join_axes=[self.data_run.index])
77
78
79     ##PnL
80     price=0.
81     for bar in range(0, len(self.final_dataframe)):

```

```

82         if self.final_dataframe['Trade'][bar] == 1:
83             price = self.get_trade_price(bar, self.final_dataframe['Units'][bar])
84         elif self.final_dataframe['Trade'][bar] == -1:
85             self.final_dataframe['PnL'].ix[bar] = self.get_trade_price(bar, self.final_dataframe[
            'Units'][bar]) - price
86
87         self.trade_stats(bar, self.final_dataframe)
88
89 if __name__ == "__main__":
90     tickerList = ["HDFCBANK", "ICICIBANK", "KOTAKBANK", "ONGC", "INFY", "RELIANCE", "HDFC", "LT", "IOC
            ", "SBIN", "HINDUNILVR",
91     "MARUTI", "ITC", "TCS"]
92
93     initial_investment_amount = 10000
94
95     for ticker in tickerList:
96         sma = MovingAverageStrategy(ticker, initial_investment_amount)
97         sma.run(50, 250)
98
99     portOBJ = Portfolio()
100     portOBJ.show_portfolio_details("port_without_hmm.pkl")

```

Chapter 7

Strategy code with using HMM

The file used to create moving average strategy with using HMM is **moving_average_using_hmm.py**. If we run this file , it will run strategy on all the listed security given in the list. Also, you can declare initial amount to be invested. The portfolio object will give you the final status and also save it on disk for future use.

It first load the model which we have trained in previous chapter. It keeps adding daily returns to a list from test data. And whenever the strategy go for trade it actually first check if any regime is detected by feeding daily returns(Observable data) to HMM model. And if model got any regime the strategy would not trade and also it will square off any holdings as well.

The function to detect regime is as follow.

```
1 def regime_detection(self, daily_returns):
2     #Converting daily return list to numpy array
3     daily_returns = np.column_stack([np.array(daily_returns)])
4     hidden_state = self.hmm_model.predict(daily_returns)[-1]
5     return hidden_state
```

If the function returns 1 it will not trade and squareoff if it has any holdings. Otherwise it will work as normal strategy.

The complete code snippet for moving average strategy with using HMM is as follow.

```
1
2 import pandas as pd
3 import numpy as np
4 import pickle
5
6 from backtest import BacktestBase
7 from portfolio import Portfolio
8
9
10 train_test_split_ratio = 0.5
11
12 class MovingAverageStrategy(BacktestBase):
13     def run(self, SMA1, SMA2, model_path):
14         daily_returns = []
15         msg = 'Running SMA strategy for %s |SMA1=%d |SMA2=%d |ftc=%f|ptc=%f'
16         msg = msg%(self.ticker, SMA1, SMA2, self.ftc, self.ptc)
17         self.position = 0
18         self.amount = self.initial_amount
19         self.trades = 0
20
21         self.hmm_model = pickle.load(open(model_path, "rb"))
```

```

22
23
24 #Data Preparation
25 self.data_run = self.data.copy()
26 self.data_run['SMA1'] = self.data_run['Close'].rolling(SMA1).mean()
27 self.data_run['SMA2'] = self.data_run['Close'].rolling(SMA2).mean()
28 self.data_run["Returns"] = self.data_run["Close"].pct_change()
29 self.data_run.dropna(inplace=True)
30
31
32 ##Split data into training and test data
33 training_data_len = int(len(self.data_run)*train_test_split_ratio) ##TO get training
data
34 end_date = self.data_run.index[training_data_len]
35 start_date = self.data_run.index[training_data_len+1]
36 training_df = self.data_run[:end_date]
37 test_df = self.data_run[start_date:]
38
39 #Running for test data only
40 self.data_run = test_df
41
42
43 # print self.data_run.head()
44
45 #####Signals
46 Signals = pd.DataFrame(index=self.data_run.index)
47 Signals["PnL"] = 0
48 Signals["Trade"] = 0
49 Signals["Units"] = 0
50
51 for bar in range(0,len(self.data_run)):
52     ##Storing observable varibale.In our case it is daily returns.

```

```

53     daily_returns.append(self.data_run[ 'Returns' ].ix[bar])
54
55     regime = self.regime_detection(daily_returns)
56
57     if self.position == 0:
58         if regime == 1:
59             pass
60         else:
61             if self.data_run[ 'SMA1' ].ix[bar] > self.data_run[ 'SMA2' ].ix[bar]:
62                 self.place_buy_order(bar, amount=self.amount)
63                 self.position = 1 #Take Position
64                 Signals["Trade"].ix[bar] = 1
65                 Signals["Units"].ix[bar] = int(self.units)
66             else:
67                 Signals["Trade"].ix[bar] = 0
68
69     elif self.position == 1:
70         if regime == 1:
71             print "Sell Without condition"
72             Signals["Units"].ix[bar] = int(self.units)
73             self.place_sell_order(bar, units=self.units)
74             self.position = 0 #Market nuetral
75             Signals["Trade"].ix[bar] = -1
76         else:
77             if self.data_run[ 'SMA1' ].ix[bar] < self.data_run[ 'SMA2' ].ix[bar]:
78                 Signals["Units"].ix[bar] = int(self.units)
79                 self.place_sell_order(bar, units=self.units)
80                 self.position = 0 #Market nuetral
81                 Signals["Trade"].ix[bar] = -1
82             else:
83                 Signals["Trade"].ix[bar] = 0
84

```

```

85
86
87
88     ##Squaring off if holds any stock without checking any condition
89     if self.position == 1:
90         Signals["Units"].ix[bar] = int(self.units)
91         self.place_sell_order(bar, units=self.units)
92         self.position = 0    #Market neutral
93         Signals["Trade"].ix[bar] = -1
94
95
96     ##Concatenating both dataframe
97     frames = [self.data_run, Signals]
98     self.final_dataframe = pd.concat(frames, axis=1, join_axes=[self.data_run.index])
99
100
101     ##PnL
102     price=0.
103
104     for bar in range(0, len(self.final_dataframe)):
105         if self.final_dataframe['Trade'][bar] == 1:
106             price = self.get_trade_price(bar, self.final_dataframe['Units'][bar])
107         elif self.final_dataframe['Trade'][bar] == -1:
108             self.final_dataframe['PnL'].ix[bar] = self.get_trade_price(bar, self.final_dataframe[
109                 'Units'][bar]) - price
110
111     self.trade_stats(bar, self.final_dataframe)
112
113     def regime_detection(self, daily_returns):
114         #Converting daily return list to numpy array
115         daily_returns = np.column_stack([np.array(daily_returns)])

```



```

116     hidden_state = self.hmm_model.predict(daily_returns)[-1]
117     return hidden_state
118
119
120 if __name__ == "__main__":
121     tickerList = ["HDFCBANK", "ICICIBANK", "KOTAKBANK", "ONGC", "INFY", "RELIANCE", "HDFC", "LT", "IOC",
122                  ", "SBIN", "HINDUNILVR",
123                  "MARUTI", "ITC", "TCS" ]
124
125
126     initial_investment_amount = 10000
127
128     for ticker in tickerList:
129         model_path = "../Models/hmm_model_"+ticker+".pkl"
130         sma = MovingAverageStrategy(ticker, initial_investment_amount)
131         sma.run(50, 250, model_path)
132
133     portOBJ = Portfolio()
134     portOBJ.show_portfolio_details("port_with_hmm.pkl")

```

Chapter 8

Findings

The portfolio details has been saved as pickle object which can be loaded and see the result. To read the result please look at file **Read_Portfolio.py**

This file just load the result file and print it in human readable format. The code snippet is as follows.

```
1 import pandas as pd
2
3 port_without_hmm = pd.read_pickle("port_without.pkl")
4 port_with_hmm = pd.read_pickle("port_with.pkl")
5
6 print "#####Portfolio Wihtout using HMM#####"
7 print port_without_hmm
8
9 print "#####Portfolio with using HMM#####"
10 print port_with_hmm
```

The result for moving average strategy without using HMM shown in figure [8.1](#).

	Invested Amount	Final Amount	Number of Trades	PnL	Sharpe Ratio	Performance
HDFCBANK	10000	10560.77899	12	560.77899	0.612232617292	5.6077899
ICICIBANK	10000	9989.614016	18	-10.385984	0.00532111748876	-0.10385984
KOTAKBANK	10000	10308.6009	20	308.6009	0.291930663207	3.086009
ONGC	10000	9198.240765	20	-801.759235	-0.545169050789	-8.01759235
INFY	10000	9718.72285	16	-281.27715	-0.223085075074	-2.8127715
RELIANCE	10000	11642.677855	12	1642.677855	0.475795629632	16.42677855
HDFC	10000	10256.49715	16	256.49715	0.271698593302	2.5649715
LT	10000	9766.627696	18	-233.372304	-0.13980268666	-2.33372304
IOC	10000	8794.833325	22	-1205.166675	-0.602970840762	-12.05166675
SBIN	10000	10440.14991	18	440.14991	0.196555503072	4.4014991
HINDUNILVR	10000	10645.769665	18	645.769665	0.259279075127	6.45769665
MARUTI	10000	10306.55655	14	306.55655	0.318665447398	3.0655655
ITC	10000	9957.459135	20	-42.540865	-0.0117827870183	-0.42540865
TCS	10000	10168.23993	16	168.23993	0.160543912718	1.6823993

Figure 8.1: Result without HMM.

The result for moving average strategy with using HMM shown in figure 8.2.

	Invested Amount	Final Amount	Number of Trades	PnL	Sharpe Ratio	Performance
HDFCBANK	10000	10539.82183	40	539.82183	0.433746837408	5.3982183
ICICIBANK	10000	9987.842031	28	-12.157969	0.0104600380918	-0.12157969
KOTAKBANK	10000	10153.75595	54	153.75595	0.137388576591	1.5375595
ONGC	10000	9226.209245	94	-773.790755	-0.326931551136	-7.73790755
INFY	10000	9814.2199	70	-185.7801	-0.0600898110903	-1.857801
RELIANCE	10000	11523.737215	24	1523.737215	0.474966781913	15.23737215
HDFC	10000	10058.3863	28	58.3863	0.156290783519	0.583863
LT	10000	9990.057544	24	-9.942456	0.0204548447429	-0.09942456
IOC	10000	8779.5593	30	-1220.4407	-0.55453045179	-12.204407
SBIN	10000	9939.5073	26	-60.4927	-0.179452317552	-0.604927
HINDUNILVR	10000	10457.693785	78	457.693785	0.271372324115	4.57693785
MARUTI	10000	10307.042875	36	307.042875	0.308808102866	3.07042875
ITC	10000	9829.2418	36	-170.7582	-0.0365560054164	-1.707582
TCS	10000	10202.795155	58	202.795155	0.13035605233	2.02795155

Figure 8.2: Result with HMM.

We can see that there are bunch of security which has performed well without using HMM such as **KOTAKBANK**, **HINDUNILVR** and **RELIANCE**. But our profit is not in negative. It is just that we earned little less profit.

Now if we compare the result for security **ONGC**, **INFY**, **LT**, **SBIN**. We can see that loss has been reduced sharply for strategy using HMM. In fact one of the security **TCS** which was in loss is giving us profit.

If you see the overall picture, we can conclude that strategy with HMM may give you little less profit but it will save you from the major losses during regime change. We can use HMM with any strategy for risk management purpose.

The result we are comparing is just for 120 days intraday data. We can actually use the large data set to see how HMM perform with your strategy for risk management.

Chapter 9

Future Work

We can do lot of thing to improve the HMM model. I have listed few of them which I will try to do it in future.

- Running different startegy with HMM such as mean reversion,pair trading.
- Changing observable varibale to std deviation,volatilty etc
- Optimizing training and test data split by using large data set.

Chapter 10

Reference

- https://en.wikipedia.org/wiki/Hidden_Markov_model
- <https://www.quantstart.com/>
- <https://www.python.org/>
- <http://hmmlearn.readthedocs.io/en/latest/tutorial.html>