Generating initial feature set

Sreejith Sreekumar

August 13, 2021

Contents

1 Features

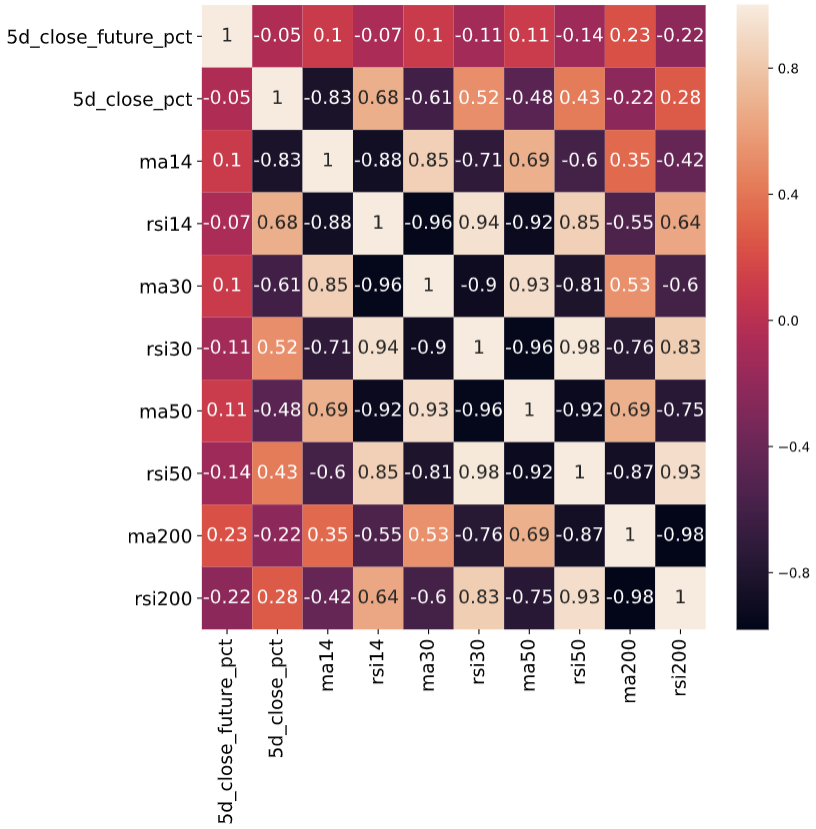
1. Simple Moving Average
    - Usual shift periods used are 14, 50 and 200
2. Relative Strength Index

$$RelativeStrength, RS = \frac{Average\ Gain\ Over\ n\ periods}{Average\ Loss\ Over\ n\ periods} \tag{1}$$

$$RSI = 100 - \frac{100}{1 + RS} \tag{2}$$

```
feature_names = ['5d_close_pct']  # a list of the feature names for later

# Create moving averages and rsi for timeperiods of 14, 30, 50, and 200
for n in [14, 30, 50, 200]:

    # Create the moving average indicator and divide by Adj_Close
    lng_df['ma' + str(n)] = talib.SMA(lng_df['Adj_Close'].values,
                              timeperiod=n) / lng_df["Adj_Close"]
    # Create the RSI indicator
    lng_df['rsi' + str(n)] = talib.RSI(lng_df['Adj_Close'].values, timeperiod=n)
```

```
    # Add rsi and moving average to the feature name list
    feature_names = feature_names + ['ma' + str(n), 'rsi' + str(n)]

print(feature_names)



# Drop all na values
lng_df = lng_df.dropna()

# Create features and targets
# use feature_names for features; '5d_close_future_pct' for targets
features = lng_df[feature_names]
targets = lng_df["5d_close_future_pct"]

# Create DataFrame from target column and feature columns
feature_and_target_cols = ['5d_close_future_pct'] + feature_names
feat_targ_df = lng_df[feature_and_target_cols]

# Calculate correlation matrix
corr = feat_targ_df.corr()
print(corr)



# Plot heatmap of correlation matrix
sns.heatmap(corr, annot= True, annot_kws = {"size": 14})
plt.yticks(rotation=0, size = 14); plt.xticks(rotation=90, size = 14)  # fix ticklabel
plt.tight_layout()  # fits plot area to the plot, "tightly"
plt.show()  # show the plot
```


5d_close_future_pct 1 -0.05 0.1 -0.07 0.1 -0.11 0.11 -0.14 0.23 -0.22
5d_close_pct -0.05 1 -0.83 0.68 -0.61 0.52 -0.48 0.43 -0.22 0.28
ma14 0.1 -0.83 1 -0.88 0.85 -0.71 0.69 -0.6 0.35 -0.42
rsi14 -0.07 0.68 -0.88 1 -0.96 0.94 -0.92 0.85 -0.55 0.64
ma30 0.1 -0.61 0.85 -0.96 1 -0.9 0.93 -0.81 0.53 -0.6
rsi30 -0.11 0.52 -0.71 0.94 -0.9 1 -0.96 0.98 -0.76 0.83
ma50 0.11 -0.48 0.69 -0.92 0.93 -0.96 1 -0.92 0.69 -0.75
rsi50 -0.14 0.43 -0.6 0.85 -0.81 0.98 -0.92 1 -0.87 0.93
ma200 0.23 -0.22 0.35 -0.55 0.53 -0.76 0.69 -0.87 1 -0.98
rsi200 -0.22 0.28 -0.42 0.64 -0.6 0.83 -0.75 0.93 -0.98 1
5d_close_future_pct
5d_close_pct
ma14
rsi14
ma30
rsi30
ma50
rsi50
ma200
rsi200
0.8
0.4
0.0
−0.4
−0.8