18.curiosity-q-learning-agent

September 29, 2021

[1]:
```
import numpy as np
import pandas as pd
import tensorflow as tf
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
```

[2]:
```
df = pd.read_csv('../dataset/GOOG-year.csv')
df.head()
```

[2]:

	Date	Open	High	Low	Close	Adj Close	Volume
0	2016-11-02	778.200012	781.650024	763.450012	768.700012	768.700012	1872400
1	2016-11-03	767.250000	769.950012	759.030029	762.130005	762.130005	1943200
2	2016-11-04	750.659973	770.359985	750.560974	762.020020	762.020020	2134800
3	2016-11-07	774.500000	785.190002	772.549988	782.520020	782.520020	1585100
4	2016-11-08	783.400024	795.632996	780.190002	790.510010	790.510010	1350800

[3]:
```
from collections import deque
import random

class Agent:

    LEARNING_RATE = 0.003
    BATCH_SIZE = 32
    LAYER_SIZE = 500
    OUTPUT_SIZE = 3
    EPSILON = 0.5
    DECAY_RATE = 0.005
    MIN_EPSILON = 0.1
    GAMMA = 0.99
```

```
    MEMORIES = deque()
    COPY = 1000
    T_COPY = 0
    MEMORY_SIZE = 300

    def __init__(self, state_size, window_size, trend, skip):
        self.state_size = state_size
        self.window_size = window_size
        self.half_window = window_size // 2
        self.trend = trend
        self.skip = skip
        tf.reset_default_graph()
        self.X = tf.placeholder(tf.float32, (None, self.state_size))
        self.Y = tf.placeholder(tf.float32, (None, self.state_size))
        self.ACTION = tf.placeholder(tf.float32, (None))
        self.REWARD = tf.placeholder(tf.float32, (None))
        self.batch_size = tf.shape(self.ACTION)[0]

        with tf.variable_scope('curiosity_model'):
            action = tf.reshape(self.ACTION, (-1,1))
            state_action = tf.concat([self.X, action], axis=1)
            save_state = tf.identity(self.Y)

            feed = tf.layers.dense(state_action, 32, activation=tf.nn.relu)
            self.curiosity_logits = tf.layers.dense(feed, self.state_size)
            self.curiosity_cost = tf.reduce_sum(tf.square(save_state - self.
↪curiosity_logits), axis=1)

            self.curiosity_optimizer = tf.train.RMSPropOptimizer(self.
↪LEARNING_RATE)\
            .minimize(tf.reduce_mean(self.curiosity_cost))

        total_reward = tf.add(self.curiosity_cost, self.REWARD)

        with tf.variable_scope("q_model"):
            with tf.variable_scope("eval_net"):
                x_action = tf.layers.dense(self.X, 128, tf.nn.relu)
                self.logits = tf.layers.dense(x_action, self.OUTPUT_SIZE)

            with tf.variable_scope("target_net"):
                y_action = tf.layers.dense(self.Y, 128, tf.nn.relu)
                y_q = tf.layers.dense(y_action, self.OUTPUT_SIZE)

            q_target = total_reward + self.GAMMA * tf.reduce_max(y_q, axis=1)
            action = tf.cast(self.ACTION, tf.int32)
            action_indices = tf.stack([tf.range(self.batch_size, dtype=tf.
↪int32), action], axis=1)
```

```
            q = tf.gather_nd(params=self.logits, indices=action_indices)
            self.cost = tf.losses.mean_squared_error(labels=q_target,␣
↪predictions=q)
            self.optimizer = tf.train.RMSPropOptimizer(self.LEARNING_RATE).
↪minimize(
            self.cost, var_list=tf.get_collection(tf.GraphKeys.
↪TRAINABLE_VARIABLES, "q_model/eval_net"))

        t_params = tf.get_collection(tf.GraphKeys.GLOBAL_VARIABLES,␣
↪scope='q_model/target_net')
        e_params = tf.get_collection(tf.GraphKeys.GLOBAL_VARIABLES,␣
↪scope='q_model/eval_net')
        self.target_replace_op = [tf.assign(t, e) for t, e in zip(t_params,␣
↪e_params)]

        self.sess = tf.InteractiveSession()
        self.sess.run(tf.global_variables_initializer())

    def _memorize(self, state, action, reward, new_state, done):
        self.MEMORIES.append((state, action, reward, new_state, done))
        if len(self.MEMORIES) > self.MEMORY_SIZE:
            self.MEMORIES.popleft()

    def get_state(self, t):
        window_size = self.window_size + 1
        d = t - window_size + 1
        block = self.trend[d : t + 1] if d >= 0 else -d * [self.trend[0]] +␣
↪self.trend[0 : t + 1]
        res = []
        for i in range(window_size - 1):
            res.append(block[i + 1] - block[i])
        return np.array(res)

    def predict(self, inputs):
        return self.sess.run(self.logits, feed_dict={self.X:inputs})

    def get_predicted_action(self, sequence):
        prediction = self.predict(np.array(sequence))[0]
        return np.argmax(prediction)

    def _select_action(self, state):
        if np.random.rand() < self.EPSILON:
            action = np.random.randint(self.OUTPUT_SIZE)
        else:
            action = self.get_predicted_action([state])
        return action
```

```
    def _construct_memories(self, replay):
        states = np.array([a[0] for a in replay])
        actions = np.array([a[1] for a in replay])
        rewards = np.array([a[2] for a in replay])
        new_states = np.array([a[3] for a in replay])
        if (self.T_COPY + 1) % self.COPY == 0:
            self.sess.run(self.target_replace_op)

        cost, _ = self.sess.run([self.cost, self.optimizer], feed_dict = {
            self.X: states, self.Y: new_states, self.ACTION: actions, self.
↪REWARD: rewards
        })

        if (self.T_COPY + 1) % self.COPY == 0:
            self.sess.run(self.curiosity_optimizer, feed_dict = {
                self.X: states, self.Y: new_states, self.ACTION: actions, self.
↪REWARD: rewards
            })
        return cost

    def buy(self, initial_money):
        starting_money = initial_money
        states_sell = []
        states_buy = []
        inventory = []
        state = self.get_state(0)
        for t in range(0, len(self.trend) - 1, self.skip):
            action = self._select_action(state)
            next_state = self.get_state(t + 1)

            if action == 1 and initial_money >= self.trend[t]:
                inventory.append(self.trend[t])
                initial_money -= self.trend[t]
                states_buy.append(t)
                print('day %d: buy 1 unit at price %f, total balance %f'% (t,␣
↪self.trend[t], initial_money))

            elif action == 2 and len(inventory):
                bought_price = inventory.pop(0)
                initial_money += self.trend[t]
                states_sell.append(t)
                try:
                    invest = ((close[t] - bought_price) / bought_price) * 100
                except:
                    invest = 0
                print(
```

```
                    'day %d, sell 1 unit at price %f, investment %f %%, total␣
↪balance %f,'
                    % (t, close[t], invest, initial_money)
                )

            state = next_state
        invest = ((initial_money - starting_money) / starting_money) * 100
        total_gains = initial_money - starting_money
        return states_buy, states_sell, total_gains, invest

    def train(self, iterations, checkpoint, initial_money):
        for i in range(iterations):
            total_profit = 0
            inventory = []
            state = self.get_state(0)
            starting_money = initial_money
            for t in range(0, len(self.trend) - 1, self.skip):

                action = self._select_action(state)
                next_state = self.get_state(t + 1)

                if action == 1 and starting_money >= self.trend[t]:
                    inventory.append(self.trend[t])
                    starting_money -= self.trend[t]

                elif action == 2 and len(inventory) > 0:
                    bought_price = inventory.pop(0)
                    total_profit += self.trend[t] - bought_price
                    starting_money += self.trend[t]

                invest = ((starting_money - initial_money) / initial_money)

                self._memorize(state, action, invest, next_state,␣
↪starting_money < initial_money)
                batch_size = min(len(self.MEMORIES), self.BATCH_SIZE)
                state = next_state
                replay = random.sample(self.MEMORIES, batch_size)
                cost = self._construct_memories(replay)
                self.T_COPY += 1
                self.EPSILON = self.MIN_EPSILON + (1.0 - self.MIN_EPSILON) * np.
↪exp(-self.DECAY_RATE * i)
            if (i+1) % checkpoint == 0:
                print('epoch: %d, total rewards: %f.3, cost: %f, total money:␣
↪%f'%(i + 1, total_profit, cost,
                                                                            ␣
↪  starting_money))
```

[4]:
```
close = df.Close.values.tolist()
initial_money = 10000
window_size = 30
skip = 1
batch_size = 32
agent = Agent(state_size = window_size,
              window_size = window_size,
              trend = close,
              skip = skip)
agent.train(iterations = 200, checkpoint = 10, initial_money = initial_money)
```

```
epoch: 10, total rewards: 2349.819823.3, cost: 69092.625000, total money:
12349.819823
epoch: 20, total rewards: 648.444882.3, cost: 4775652.000000, total money:
6742.654903
epoch: 30, total rewards: 1543.784977.3, cost: 26533.583984, total money:
7642.034916
epoch: 40, total rewards: 1360.930418.3, cost: 871420.750000, total money:
695.580380
epoch: 50, total rewards: 2233.069826.3, cost: 228718.296875, total money:
6354.209779
epoch: 60, total rewards: 1573.414983.3, cost: 407432.843750, total money:
8625.614995
epoch: 70, total rewards: -7.114931.3, cost: 32132.660156, total money:
5021.405088
epoch: 80, total rewards: 798.045042.3, cost: 435778.562500, total money:
9780.935057
epoch: 90, total rewards: 575.719967.3, cost: 72847.468750, total money:
9559.079952
epoch: 100, total rewards: 338.655157.3, cost: 379671.968750, total money:
820.245184
epoch: 110, total rewards: 277.220155.3, cost: 391019.375000, total money:
3452.330140
epoch: 120, total rewards: 370.379826.3, cost: 429969.843750, total money:
7361.909793
epoch: 130, total rewards: 441.860107.3, cost: 2082513.625000, total money:
2538.970093
epoch: 140, total rewards: 709.099850.3, cost: 558315.562500, total money:
130.919796
epoch: 150, total rewards: 159.675106.3, cost: 2904243.000000, total money:
481.725093
epoch: 160, total rewards: 581.489981.3, cost: 1408646.250000, total money:
5631.309988
epoch: 170, total rewards: 1768.579776.3, cost: 1693698.250000, total money:
15.189760
epoch: 180, total rewards: 952.280210.3, cost: 1472623.250000, total money:
8990.750181
```

```
epoch: 190, total rewards: 1418.655145.3, cost: 25627934.000000, total money:
3706.275139
epoch: 200, total rewards: 272.595214.3, cost: 922414.500000, total money:
9255.485229
```

[9]:
```
states_buy, states_sell, total_gains, invest = agent.buy(initial_money =␣
↪initial_money)
```

```
day 7: buy 1 unit at price 754.020020, total balance 9245.979980
day 8: buy 1 unit at price 736.080017, total balance 8509.899963
day 9: buy 1 unit at price 758.489990, total balance 7751.409973
day 10: buy 1 unit at price 764.479980, total balance 6986.929993
day 11: buy 1 unit at price 771.229980, total balance 6215.700013
day 15, sell 1 unit at price 760.989990, investment 0.924375 %, total balance
6976.690003,
day 18: buy 1 unit at price 770.840027, total balance 6205.849976
day 19: buy 1 unit at price 758.039978, total balance 5447.809998
day 22: buy 1 unit at price 762.520020, total balance 4685.289978
day 23: buy 1 unit at price 759.109985, total balance 3926.179993
day 24, sell 1 unit at price 771.190002, investment 4.769860 %, total balance
4697.369995,
day 27: buy 1 unit at price 789.270020, total balance 3908.099975
day 28, sell 1 unit at price 796.099976, investment 4.958534 %, total balance
4704.199951,
day 29: buy 1 unit at price 797.070007, total balance 3907.129944
day 30, sell 1 unit at price 797.849976, investment 4.365058 %, total balance
4704.979920,
day 31: buy 1 unit at price 790.799988, total balance 3914.179932
day 32, sell 1 unit at price 794.200012, investment 2.978363 %, total balance
4708.379944,
day 33: buy 1 unit at price 796.419983, total balance 3911.959961
day 35: buy 1 unit at price 791.260010, total balance 3120.699951
day 36, sell 1 unit at price 789.909973, investment 2.473917 %, total balance
3910.609924,
day 37: buy 1 unit at price 791.549988, total balance 3119.059936
day 40, sell 1 unit at price 771.820007, investment 1.817850 %, total balance
3890.879943,
day 41, sell 1 unit at price 786.140015, investment 3.097623 %, total balance
4677.019958,
day 42, sell 1 unit at price 786.900024, investment 3.660871 %, total balance
5463.919982,
day 43, sell 1 unit at price 794.020020, investment 0.601822 %, total balance
6257.940002,
day 44: buy 1 unit at price 806.150024, total balance 5451.789978
day 45: buy 1 unit at price 806.650024, total balance 4645.139954
day 47: buy 1 unit at price 807.909973, total balance 3837.229981
day 48, sell 1 unit at price 806.359985, investment 1.165516 %, total balance
4643.589966,
```

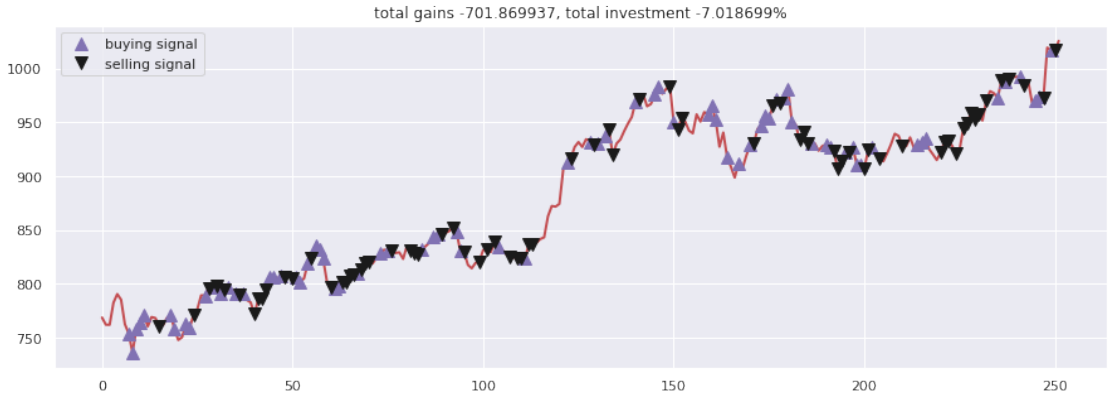
day 50, sell 1 unit at price 804.609985, investment 1.746332 %, total balance 5448.199951,
day 51: buy 1 unit at price 806.070007, total balance 4642.129944
day 52: buy 1 unit at price 802.174988, total balance 3839.954956
day 54: buy 1 unit at price 819.309998, total balance 3020.644958
day 55, sell 1 unit at price 823.869995, investment 3.446675 %, total balance 3844.514953,
day 56: buy 1 unit at price 835.669983, total balance 3008.844970
day 57: buy 1 unit at price 832.150024, total balance 2176.694946
day 58: buy 1 unit at price 823.309998, total balance 1353.384948
day 60, sell 1 unit at price 796.789978, investment 0.698881 %, total balance 2150.174926,
day 61: buy 1 unit at price 795.695007, total balance 1354.479919
day 62: buy 1 unit at price 798.530029, total balance 555.949890
day 63, sell 1 unit at price 801.489990, investment 1.255764 %, total balance 1357.439880,
day 64, sell 1 unit at price 801.340027, investment -0.596663 %, total balance 2158.779907,
day 65, sell 1 unit at price 806.969971, investment 0.039664 %, total balance 2965.749878,
day 66, sell 1 unit at price 808.380005, investment 0.058179 %, total balance 3774.129883,
day 67: buy 1 unit at price 809.559998, total balance 2964.569885
day 68, sell 1 unit at price 813.669983, investment 0.942843 %, total balance 3778.239868,
day 69, sell 1 unit at price 819.239990, investment 2.127342 %, total balance 4597.479858,
day 70, sell 1 unit at price 820.450012, investment 0.139143 %, total balance 5417.929870,
day 73: buy 1 unit at price 828.070007, total balance 4589.859863
day 75: buy 1 unit at price 830.760010, total balance 3759.099853
day 76, sell 1 unit at price 831.330017, investment -0.519340 %, total balance 4590.429870,
day 81, sell 1 unit at price 830.630005, investment -0.182662 %, total balance 5421.059875,
day 82, sell 1 unit at price 829.080017, investment 0.700832 %, total balance 6250.139892,
day 83, sell 1 unit at price 827.780029, investment 4.032327 %, total balance 7077.919921,
day 84: buy 1 unit at price 831.909973, total balance 6246.009948
day 87: buy 1 unit at price 843.250000, total balance 5402.759948
day 88: buy 1 unit at price 845.539978, total balance 4557.219970
day 89, sell 1 unit at price 845.619995, investment 5.897081 %, total balance 5402.839965,
day 92, sell 1 unit at price 852.119995, investment 5.257176 %, total balance 6254.959960,
day 93: buy 1 unit at price 848.400024, total balance 5406.559936
day 94: buy 1 unit at price 830.460022, total balance 4576.099914

day 95, sell 1 unit at price 829.590027, investment 0.183562 %, total balance 5405.689941,
day 99, sell 1 unit at price 820.919983, investment -1.184461 %, total balance 6226.609924,
day 101, sell 1 unit at price 831.500000, investment -0.049281 %, total balance 7058.109924,
day 103, sell 1 unit at price 838.549988, investment -0.557369 %, total balance 7896.659912,
day 104: buy 1 unit at price 834.570007, total balance 7062.089905
day 107, sell 1 unit at price 824.669983, investment -2.468245 %, total balance 7886.759888,
day 109, sell 1 unit at price 823.349976, investment -2.952622 %, total balance 8710.109864,
day 110, sell 1 unit at price 824.320007, investment -0.739351 %, total balance 9534.429871,
day 111: buy 1 unit at price 823.559998, total balance 8710.869873
day 112, sell 1 unit at price 837.169983, investment 0.311535 %, total balance 9548.039856,
day 113, sell 1 unit at price 836.820007, investment 1.610084 %, total balance 10384.859863,
day 122: buy 1 unit at price 912.570007, total balance 9472.289856
day 123, sell 1 unit at price 916.440002, investment 0.424077 %, total balance 10388.729858,
day 128: buy 1 unit at price 932.169983, total balance 9456.559875
day 129, sell 1 unit at price 928.780029, investment -0.363663 %, total balance 10385.339904,
day 130: buy 1 unit at price 930.599976, total balance 9454.739928
day 132: buy 1 unit at price 937.080017, total balance 8517.659911
day 133, sell 1 unit at price 943.000000, investment 1.332476 %, total balance 9460.659911,
day 134, sell 1 unit at price 919.619995, investment -1.863237 %, total balance 10380.279906,
day 140: buy 1 unit at price 969.539978, total balance 9410.739928
day 141, sell 1 unit at price 971.469971, investment 0.199063 %, total balance 10382.209899,
day 145: buy 1 unit at price 975.599976, total balance 9406.609923
day 146: buy 1 unit at price 983.679993, total balance 8422.929930
day 149, sell 1 unit at price 983.409973, investment 0.800533 %, total balance 9406.339903,
day 150: buy 1 unit at price 949.830017, total balance 8456.509886
day 151, sell 1 unit at price 942.900024, investment -4.145654 %, total balance 9399.409910,
day 152, sell 1 unit at price 953.400024, investment 0.375857 %, total balance 10352.809934,
day 159: buy 1 unit at price 957.090027, total balance 9395.719907
day 160: buy 1 unit at price 965.590027, total balance 8430.129880
day 161: buy 1 unit at price 952.270020, total balance 7477.859860
day 164: buy 1 unit at price 917.789978, total balance 6560.069882

```
day 167: buy 1 unit at price 911.710022, total balance 5648.359860
day 170: buy 1 unit at price 928.799988, total balance 4719.559872
day 171, sell 1 unit at price 930.090027, investment -2.821051 %, total balance
5649.649899,
day 173: buy 1 unit at price 947.159973, total balance 4702.489926
day 174: buy 1 unit at price 955.989990, total balance 3746.499936
day 175: buy 1 unit at price 953.419983, total balance 2793.079953
day 176, sell 1 unit at price 965.400024, investment -0.019677 %, total balance
3758.479977,
day 177: buy 1 unit at price 970.890015, total balance 2787.589962
day 178, sell 1 unit at price 968.150024, investment 1.667595 %, total balance
3755.739986,
day 179: buy 1 unit at price 972.919983, total balance 2782.820003
day 180: buy 1 unit at price 980.340027, total balance 1802.479976
day 181: buy 1 unit at price 950.700012, total balance 851.779964
day 183, sell 1 unit at price 934.090027, investment 1.776011 %, total balance
1785.869991,
day 184, sell 1 unit at price 941.530029, investment 3.270778 %, total balance
2727.400020,
day 185, sell 1 unit at price 930.500000, investment 0.183033 %, total balance
3657.900020,
day 186: buy 1 unit at price 930.830017, total balance 2727.070003
day 190: buy 1 unit at price 929.359985, total balance 1797.710018
day 191: buy 1 unit at price 926.789978, total balance 870.920040
day 192, sell 1 unit at price 922.900024, investment -2.561336 %, total balance
1793.820064,
day 193, sell 1 unit at price 907.239990, investment -5.099426 %, total balance
2701.060054,
day 194, sell 1 unit at price 914.390015, investment -4.093681 %, total balance
3615.450069,
day 195: buy 1 unit at price 922.669983, total balance 2692.780086
day 196, sell 1 unit at price 922.219971, investment -5.012931 %, total balance
3615.000057,
day 197: buy 1 unit at price 926.960022, total balance 2688.040035
day 198: buy 1 unit at price 910.979980, total balance 1777.060055
day 199: buy 1 unit at price 910.669983, total balance 866.390072
day 200, sell 1 unit at price 906.659973, investment -6.810427 %, total balance
1773.050045,
day 201, sell 1 unit at price 924.690002, investment -5.676604 %, total balance
2697.740047,
day 202: buy 1 unit at price 927.000000, total balance 1770.740047
day 204, sell 1 unit at price 915.890015, investment -3.661512 %, total balance
2686.630062,
day 210, sell 1 unit at price 928.450012, investment -0.255686 %, total balance
3615.080074,
day 214: buy 1 unit at price 929.080017, total balance 2686.000057
day 215: buy 1 unit at price 932.070007, total balance 1753.930050
day 216: buy 1 unit at price 935.090027, total balance 818.840023
```

```
day 220, sell 1 unit at price 921.809998, investment -0.812386 %, total balance
1740.650021,
day 221, sell 1 unit at price 931.580017, investment 0.516842 %, total balance
2672.230038,
day 222, sell 1 unit at price 932.450012, investment 1.059970 %, total balance
3604.680050,
day 224, sell 1 unit at price 920.969971, investment -0.646204 %, total balance
4525.650021,
day 226, sell 1 unit at price 944.489990, investment 3.678457 %, total balance
5470.140011,
day 227, sell 1 unit at price 949.500000, investment 4.263896 %, total balance
6419.640011,
day 228, sell 1 unit at price 959.109985, investment 3.463860 %, total balance
7378.749996,
day 229, sell 1 unit at price 953.270020, investment 2.603651 %, total balance
8332.020016,
day 230, sell 1 unit at price 957.789978, investment 2.759446 %, total balance
9289.809994,
day 232, sell 1 unit at price 969.960022, investment 3.729052 %, total balance
10259.770016,
day 235: buy 1 unit at price 972.599976, total balance 9287.170040
day 236, sell 1 unit at price 989.250000, investment 1.711909 %, total balance
10276.420040,
day 237: buy 1 unit at price 987.830017, total balance 9288.590023
day 238, sell 1 unit at price 989.679993, investment 0.187277 %, total balance
10278.270016,
day 241: buy 1 unit at price 992.809998, total balance 9285.460018
day 242, sell 1 unit at price 984.450012, investment -0.842053 %, total balance
10269.910030,
day 245: buy 1 unit at price 970.539978, total balance 9299.370052
day 246: buy 1 unit at price 973.330017, total balance 8326.040035
day 247, sell 1 unit at price 972.559998, investment 0.208134 %, total balance
9298.600033,
day 249: buy 1 unit at price 1017.109985, total balance 8281.490048
day 250, sell 1 unit at price 1016.640015, investment 4.449672 %, total balance
9298.130063,
```

[10]:
```
fig = plt.figure(figsize = (15,5))
plt.plot(close, color='r', lw=2.)
plt.plot(close, '^', markersize=10, color='m', label = 'buying signal',␣
 ↪markevery = states_buy)
plt.plot(close, 'v', markersize=10, color='k', label = 'selling signal',␣
 ↪markevery = states_sell)
plt.title('total gains %f, total investment %f%%'%(total_gains, invest))
plt.legend()
plt.show()
```


total gains -701.869937, total investment -7.018699%
buying signal
selling signal
1000
950
900
850
800
750
0
50
100
150
200
250

[ ]: