7.double-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  \
0  2016-11-02  778.200012  781.650024  763.450012  768.700012  768.700012
1  2016-11-03  767.250000  769.950012  759.030029  762.130005  762.130005
2  2016-11-04  750.659973  770.359985  750.560974  762.020020  762.020020
3  2016-11-07  774.500000  785.190002  772.549988  782.520020  782.520020
4  2016-11-08  783.400024  795.632996  780.190002  790.510010  790.510010

    Volume
0  1872400
1  1943200
2  2134800
3  1585100
4  1350800
```

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

class Model:
    def __init__(self, input_size, output_size, layer_size, learning_rate):
        self.X = tf.placeholder(tf.float32, (None, input_size))
        self.Y = tf.placeholder(tf.float32, (None, output_size))
        feed_forward = tf.layers.dense(self.X, layer_size, activation = tf.nn.
 ↪relu)
        self.logits = tf.layers.dense(feed_forward, output_size)
        self.cost = tf.reduce_sum(tf.square(self.Y - self.logits))
        self.optimizer = tf.train.AdamOptimizer(learning_rate = learning_rate).
 ↪minimize(self.cost)
```

```
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.model = Model(self.state_size, self.OUTPUT_SIZE, self.LAYER_SIZE,␣
↪self.LEARNING_RATE)
        self.model_negative = Model(self.state_size, self.OUTPUT_SIZE, self.
↪LAYER_SIZE, self.LEARNING_RATE)
        self.sess = tf.InteractiveSession()
        self.sess.run(tf.global_variables_initializer())
        self.trainable = tf.trainable_variables()

    def _assign(self):
        for i in range(len(self.trainable)//2):
            assign_op = self.trainable[i+len(self.trainable)//2].assign(self.
↪trainable[i])
            self.sess.run(assign_op)

    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 _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])
        new_states = np.array([a[3] for a in replay])
        Q = self.predict(states)
        Q_new = self.predict(new_states)
        Q_new_negative = self.sess.run(self.model_negative.logits,␣
↪feed_dict={self.model_negative.X:new_states})
        replay_size = len(replay)
        X = np.empty((replay_size, self.state_size))
        Y = np.empty((replay_size, self.OUTPUT_SIZE))
        for i in range(replay_size):
            state_r, action_r, reward_r, new_state_r, done_r = replay[i]
            target = Q[i]
            target[action_r] = reward_r
            if not done_r:
                target[action_r] += self.GAMMA * Q_new_negative[i, np.
↪argmax(Q_new[i])]
            X[i] = state_r
            Y[i] = target
        return X, Y

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

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

    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 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):
                if (self.T_COPY + 1) % self.COPY == 0:
                    self._assign()

                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)
                replay = random.sample(self.MEMORIES, batch_size)
                state = next_state
                X, Y = self._construct_memories(replay)

                cost, _ = self.sess.run([self.model.cost, self.model.optimizer],
                                        feed_dict={self.model.X: X, self.model.
↪Y:Y})
                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: 1241.885127.3, cost: 1.110860, total money:
1744.875178
epoch: 20, total rewards: 89.105106.3, cost: 0.649060, total money: 8097.275088
epoch: 30, total rewards: 719.079470.3, cost: 0.823131, total money: 9699.809450
epoch: 40, total rewards: 684.040043.3, cost: 1.931746, total money: 134.750004
epoch: 50, total rewards: 1744.829771.3, cost: 0.895153, total money:
11744.829771
epoch: 60, total rewards: 149.195010.3, cost: 1.097174, total money: 5196.854982
epoch: 70, total rewards: 1389.289786.3, cost: 0.860031, total money:
9399.319754
```

```
epoch: 80, total rewards: 529.019898.3, cost: 0.305593, total money:
10529.019898
epoch: 90, total rewards: 1285.264893.3, cost: 1.882383, total money:
9251.514893
epoch: 100, total rewards: 409.474970.3, cost: 0.146280, total money: 551.414972
epoch: 110, total rewards: 1074.725155.3, cost: 0.661549, total money:
2231.475154
epoch: 120, total rewards: 1713.854676.3, cost: 1.219318, total money:
11713.854676
epoch: 130, total rewards: 871.945621.3, cost: 1.460638, total money:
8947.665652
epoch: 140, total rewards: 1564.314818.3, cost: 1.133385, total money:
2767.354796
epoch: 150, total rewards: 855.729796.3, cost: 1.886093, total money:
10855.729796
epoch: 160, total rewards: 302.970157.3, cost: 0.642825, total money:
6320.700137
epoch: 170, total rewards: 512.139521.3, cost: 3.411159, total money:
1801.649470
epoch: 180, total rewards: 769.354739.3, cost: 0.379282, total money:
10769.354739
epoch: 190, total rewards: 332.274720.3, cost: 1.111366, total money:
10332.274720
epoch: 200, total rewards: 395.419923.3, cost: 0.270106, total money:
5401.389893
```

[5]:
```
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 9, sell 1 unit at price 758.489990, investment 0.592818 %, total balance
10004.469970,
day 10: buy 1 unit at price 764.479980, total balance 9239.989990
day 11: buy 1 unit at price 771.229980, total balance 8468.760010
day 12, sell 1 unit at price 760.539978, investment -0.515383 %, total balance
9229.299988,
day 13, sell 1 unit at price 769.200012, investment -0.263212 %, total balance
9998.500000,
day 17: buy 1 unit at price 768.239990, total balance 9230.260010
day 19, sell 1 unit at price 758.039978, investment -1.327712 %, total balance
9988.299988,
day 21: buy 1 unit at price 750.500000, total balance 9237.799988
day 22, sell 1 unit at price 762.520020, investment 1.601602 %, total balance
10000.320008,
day 26: buy 1 unit at price 789.289978, total balance 9211.030030
day 27, sell 1 unit at price 789.270020, investment -0.002529 %, total balance
10000.300050,
day 30: buy 1 unit at price 797.849976, total balance 9202.450074
```

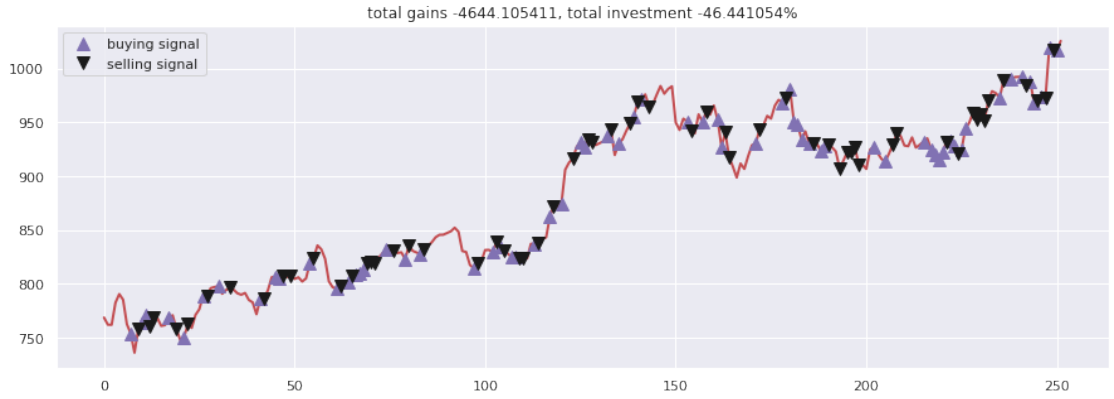
```
day 33, sell 1 unit at price 796.419983, investment -0.179231 %, total balance
9998.870057,
day 41: buy 1 unit at price 786.140015, total balance 9212.730042
day 42, sell 1 unit at price 786.900024, investment 0.096676 %, total balance
9999.630066,
day 45: buy 1 unit at price 806.650024, total balance 9192.980042
day 46: buy 1 unit at price 804.789978, total balance 8388.190064
day 47, sell 1 unit at price 807.909973, investment 0.156195 %, total balance
9196.100037,
day 49, sell 1 unit at price 807.880005, investment 0.383954 %, total balance
10003.980042,
day 54: buy 1 unit at price 819.309998, total balance 9184.670044
day 55, sell 1 unit at price 823.869995, investment 0.556566 %, total balance
10008.540039,
day 61: buy 1 unit at price 795.695007, total balance 9212.845032
day 62, sell 1 unit at price 798.530029, investment 0.356295 %, total balance
10011.375061,
day 64: buy 1 unit at price 801.340027, total balance 9210.035034
day 65, sell 1 unit at price 806.969971, investment 0.702566 %, total balance
10017.005005,
day 66: buy 1 unit at price 808.380005, total balance 9208.625000
day 67: buy 1 unit at price 809.559998, total balance 8399.065002
day 68: buy 1 unit at price 813.669983, total balance 7585.395019
day 69, sell 1 unit at price 819.239990, investment 1.343426 %, total balance
8404.635009,
day 70, sell 1 unit at price 820.450012, investment 1.345177 %, total balance
9225.085021,
day 71, sell 1 unit at price 818.979980, investment 0.652598 %, total balance
10044.065001,
day 74: buy 1 unit at price 831.659973, total balance 9212.405028
day 76, sell 1 unit at price 831.330017, investment -0.039674 %, total balance
10043.735045,
day 79: buy 1 unit at price 823.210022, total balance 9220.525023
day 80, sell 1 unit at price 835.239990, investment 1.461349 %, total balance
10055.765013,
day 83: buy 1 unit at price 827.780029, total balance 9227.984984
day 84, sell 1 unit at price 831.909973, investment 0.498918 %, total balance
10059.894957,
day 97: buy 1 unit at price 814.429993, total balance 9245.464964
day 98, sell 1 unit at price 819.510010, investment 0.623751 %, total balance
10064.974974,
day 102: buy 1 unit at price 829.559998, total balance 9235.414976
day 103, sell 1 unit at price 838.549988, investment 1.083706 %, total balance
10073.964964,
day 104: buy 1 unit at price 834.570007, total balance 9239.394957
day 105, sell 1 unit at price 831.409973, investment -0.378642 %, total balance
10070.804930,
day 107: buy 1 unit at price 824.669983, total balance 9246.134947
```

```
day 108: buy 1 unit at price 824.729980, total balance 8421.404967
day 109, sell 1 unit at price 823.349976, investment -0.160065 %, total balance
9244.754943,
day 110, sell 1 unit at price 824.320007, investment -0.049710 %, total balance
10069.074950,
day 113: buy 1 unit at price 836.820007, total balance 9232.254943
day 114, sell 1 unit at price 838.210022, investment 0.166107 %, total balance
10070.464965,
day 117: buy 1 unit at price 862.760010, total balance 9207.704955
day 118, sell 1 unit at price 872.299988, investment 1.105751 %, total balance
10080.004943,
day 120: buy 1 unit at price 874.250000, total balance 9205.754943
day 123, sell 1 unit at price 916.440002, investment 4.825851 %, total balance
10122.194945,
day 125: buy 1 unit at price 931.659973, total balance 9190.534972
day 126: buy 1 unit at price 927.130005, total balance 8263.404967
day 127, sell 1 unit at price 934.299988, investment 0.283367 %, total balance
9197.704955,
day 128, sell 1 unit at price 932.169983, investment 0.543611 %, total balance
10129.874938,
day 132: buy 1 unit at price 937.080017, total balance 9192.794921
day 133, sell 1 unit at price 943.000000, investment 0.631748 %, total balance
10135.794921,
day 135: buy 1 unit at price 930.239990, total balance 9205.554931
day 138, sell 1 unit at price 948.820007, investment 1.997336 %, total balance
10154.374938,
day 139: buy 1 unit at price 954.960022, total balance 9199.414916
day 140, sell 1 unit at price 969.539978, investment 1.526761 %, total balance
10168.954894,
day 141: buy 1 unit at price 971.469971, total balance 9197.484923
day 143, sell 1 unit at price 964.859985, investment -0.680411 %, total balance
10162.344908,
day 153: buy 1 unit at price 950.760010, total balance 9211.584898
day 154, sell 1 unit at price 942.309998, investment -0.888764 %, total balance
10153.894896,
day 157: buy 1 unit at price 950.630005, total balance 9203.264891
day 158, sell 1 unit at price 959.450012, investment 0.927807 %, total balance
10162.714903,
day 161: buy 1 unit at price 952.270020, total balance 9210.444883
day 162: buy 1 unit at price 927.330017, total balance 8283.114866
day 163, sell 1 unit at price 940.489990, investment -1.237047 %, total balance
9223.604856,
day 164, sell 1 unit at price 917.789978, investment -1.028764 %, total balance
10141.394834,
day 171: buy 1 unit at price 930.090027, total balance 9211.304807
day 172, sell 1 unit at price 943.830017, investment 1.477275 %, total balance
10155.134824,
day 178: buy 1 unit at price 968.150024, total balance 9186.984800
```

```
day 179, sell 1 unit at price 972.919983, investment 0.492688 %, total balance
10159.904783,
day 180: buy 1 unit at price 980.340027, total balance 9179.564756
day 181: buy 1 unit at price 950.700012, total balance 8228.864744
day 182: buy 1 unit at price 947.799988, total balance 7281.064756
day 183: buy 1 unit at price 934.090027, total balance 6346.974729
day 185: buy 1 unit at price 930.500000, total balance 5416.474729
day 186, sell 1 unit at price 930.830017, investment -5.050290 %, total balance
6347.304746,
day 188: buy 1 unit at price 923.650024, total balance 5423.654722
day 189: buy 1 unit at price 927.960022, total balance 4495.694700
day 190, sell 1 unit at price 929.359985, investment -2.244665 %, total balance
5425.054685,
day 193, sell 1 unit at price 907.239990, investment -4.279384 %, total balance
6332.294675,
day 195, sell 1 unit at price 922.669983, investment -1.222585 %, total balance
7254.964658,
day 196, sell 1 unit at price 922.219971, investment -0.889847 %, total balance
8177.184629,
day 197, sell 1 unit at price 926.960022, investment 0.358361 %, total balance
9104.144651,
day 198, sell 1 unit at price 910.979980, investment -1.829825 %, total balance
10015.124631,
day 202: buy 1 unit at price 927.000000, total balance 9088.124631
day 205: buy 1 unit at price 913.809998, total balance 8174.314633
day 207, sell 1 unit at price 929.570007, investment 0.277239 %, total balance
9103.884640,
day 208, sell 1 unit at price 939.330017, investment 2.792705 %, total balance
10043.214657,
day 215: buy 1 unit at price 932.070007, total balance 9111.144650
day 217: buy 1 unit at price 925.109985, total balance 8186.034665
day 218: buy 1 unit at price 920.289978, total balance 7265.744687
day 219: buy 1 unit at price 915.000000, total balance 6350.744687
day 220: buy 1 unit at price 921.809998, total balance 5428.934689
day 221, sell 1 unit at price 931.580017, investment -0.052570 %, total balance
6360.514706,
day 222: buy 1 unit at price 932.450012, total balance 5428.064694
day 223: buy 1 unit at price 928.530029, total balance 4499.534665
day 224, sell 1 unit at price 920.969971, investment -0.447516 %, total balance
5420.504636,
day 225: buy 1 unit at price 924.859985, total balance 4495.644651
day 226: buy 1 unit at price 944.489990, total balance 3551.154661
day 228, sell 1 unit at price 959.109985, investment 4.218236 %, total balance
4510.264646,
day 229, sell 1 unit at price 953.270020, investment 4.182516 %, total balance
5463.534666,
day 230, sell 1 unit at price 957.789978, investment 3.903188 %, total balance
6421.324644,
```

```
day 231, sell 1 unit at price 951.679993, investment 2.062307 %, total balance
7373.004637,
day 232, sell 1 unit at price 969.960022, investment 4.461890 %, total balance
8342.964659,
day 235: buy 1 unit at price 972.599976, total balance 7370.364683
day 236, sell 1 unit at price 989.250000, investment 6.962137 %, total balance
8359.614683,
day 238: buy 1 unit at price 989.679993, total balance 7369.934690
day 241: buy 1 unit at price 992.809998, total balance 6377.124692
day 242, sell 1 unit at price 984.450012, investment 4.230857 %, total balance
7361.574704,
day 243: buy 1 unit at price 988.200012, total balance 6373.374692
day 244: buy 1 unit at price 968.450012, total balance 5404.924680
day 245, sell 1 unit at price 970.539978, investment -0.211803 %, total balance
6375.464658,
day 246: buy 1 unit at price 973.330017, total balance 5402.134641
day 247, sell 1 unit at price 972.559998, investment -1.729852 %, total balance
6374.694639,
day 248: buy 1 unit at price 1019.270020, total balance 5355.424619
day 249, sell 1 unit at price 1017.109985, investment 2.447597 %, total balance
6372.534604,
day 250: buy 1 unit at price 1016.640015, total balance 5355.894589
```

[6]:
```
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()
```



[ ]: