10.duel-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:
    def __init__(self, state_size, window_size, trend, skip, batch_size):
        self.state_size = state_size
        self.window_size = window_size
        self.half_window = window_size // 2
        self.trend = trend
        self.skip = skip
        self.action_size = 3
        self.batch_size = batch_size
```

```
        self.memory = deque(maxlen = 1000)
        self.inventory = []

        self.gamma = 0.95
        self.epsilon = 0.5
        self.epsilon_min = 0.01
        self.epsilon_decay = 0.999

        tf.reset_default_graph()
        self.sess = tf.InteractiveSession()
        self.X = tf.placeholder(tf.float32, [None, self.state_size])
        self.Y = tf.placeholder(tf.float32, [None, self.action_size])
        feed = tf.layers.dense(self.X, 512, activation = tf.nn.relu)
        tensor_action, tensor_validation = tf.split(feed,2,1)
        feed_action = tf.layers.dense(tensor_action, self.action_size)
        feed_validation = tf.layers.dense(tensor_validation, 1)
        self.logits = feed_validation + tf.subtract(feed_action,tf.
↪reduce_mean(feed_action,axis=1,keep_dims=True))
        self.cost = tf.reduce_mean(tf.square(self.Y - self.logits))
        self.optimizer = tf.train.GradientDescentOptimizer(1e-5).minimize(
            self.cost
        )
        self.sess.run(tf.global_variables_initializer())

    def act(self, state):
        if random.random() <= self.epsilon:
            return random.randrange(self.action_size)
        return np.argmax(
            self.sess.run(self.logits, feed_dict = {self.X: state})[0]
        )

    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 replay(self, batch_size):
        mini_batch = []
        l = len(self.memory)
        for i in range(l - batch_size, l):
            mini_batch.append(self.memory[i])
        replay_size = len(mini_batch)
```

```
        X = np.empty((replay_size, self.state_size))
        Y = np.empty((replay_size, self.action_size))
        states = np.array([a[0][0] for a in mini_batch])
        new_states = np.array([a[3][0] for a in mini_batch])
        Q = self.sess.run(self.logits, feed_dict = {self.X: states})
        Q_new = self.sess.run(self.logits, feed_dict = {self.X: new_states})
        for i in range(len(mini_batch)):
            state, action, reward, next_state, done = mini_batch[i]
            target = Q[i]
            target[action] = reward
            if not done:
                target[action] += self.gamma * np.amax(Q_new[i])
            X[i] = state
            Y[i] = target
        cost, _ = self.sess.run(
            [self.cost, self.optimizer], feed_dict = {self.X: X, self.Y: Y}
        )
        if self.epsilon > self.epsilon_min:
            self.epsilon *= self.epsilon_decay
        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.act(state)
            next_state = self.get_state(t + 1)

            if action == 1 and initial_money >= self.trend[t] and t < (len(self.
↪trend) - self.half_window):
                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.act(state)
                next_state = self.get_state(t + 1)

                if action == 1 and starting_money >= self.trend[t] and t <␣
↪(len(self.trend) - self.half_window):
                    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.memory.append((state, action, invest,
                                    next_state, starting_money < initial_money))
                state = next_state
                batch_size = min(self.batch_size, len(self.memory))
                cost = self.replay(batch_size)
            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,
              batch_size = batch_size)
agent.train(iterations = 200, checkpoint = 10, initial_money = initial_money)
```

```
WARNING:tensorflow:From <ipython-input-3-28bed545c0f8>:30: calling reduce_mean
(from tensorflow.python.ops.math_ops) with keep_dims is deprecated and will be
removed in a future version.
Instructions for updating:
keep_dims is deprecated, use keepdims instead
epoch: 10, total rewards: 231.100222.3, cost: 0.499693, total money:
10231.100222
epoch: 20, total rewards: 195.875063.3, cost: 0.324152, total money:
10195.875063
epoch: 30, total rewards: 219.615054.3, cost: 0.237771, total money:
10219.615054
epoch: 40, total rewards: 56.505131.3, cost: 0.183305, total money: 10056.505131
epoch: 50, total rewards: 190.745120.3, cost: 0.129967, total money:
10190.745120
epoch: 60, total rewards: 165.275088.3, cost: 0.134246, total money:
10165.275088
epoch: 70, total rewards: 201.795107.3, cost: 0.075016, total money:
10201.795107
epoch: 80, total rewards: 187.545045.3, cost: 0.062454, total money:
10187.545045
epoch: 90, total rewards: 206.835023.3, cost: 0.050687, total money:
10206.835023
epoch: 100, total rewards: 199.895082.3, cost: 0.041359, total money:
10199.895082
epoch: 110, total rewards: 184.405092.3, cost: 0.035289, total money:
10184.405092
epoch: 120, total rewards: 242.405092.3, cost: 0.047248, total money:
10242.405092
epoch: 130, total rewards: 148.405032.3, cost: 0.050786, total money:
10148.405032
epoch: 140, total rewards: 225.724978.3, cost: 0.021171, total money:
10225.724978
epoch: 150, total rewards: 168.344972.3, cost: 0.018388, total money:
10168.344972
epoch: 160, total rewards: 230.095034.3, cost: 0.199324, total money:
10230.095034
epoch: 170, total rewards: 206.275026.3, cost: 0.044696, total money:
10206.275026
```

```
epoch: 180, total rewards: 364.895023.3, cost: 0.016494, total money:
10364.895023
epoch: 190, total rewards: 220.664980.3, cost: 0.014381, total money:
10220.664980
epoch: 200, total rewards: 175.284975.3, cost: 0.010883, total money:
10175.284975
```

[5]:
```
states_buy, states_sell, total_gains, invest = agent.buy(initial_money =␣
↪initial_money)
```

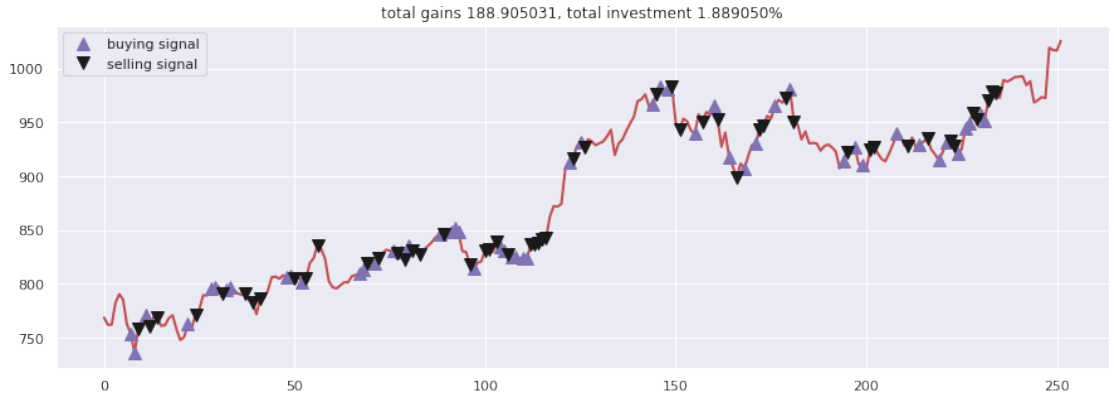
```
day 7: buy 1 unit at price 754.020020, total balance 9245.979980
day 8: buy 1 unit at price 736.080017, total balance 8509.899963
day 9, sell 1 unit at price 758.489990, investment 0.592818 %, total balance
9268.389953,
day 11: buy 1 unit at price 771.229980, total balance 8497.159973
day 12, sell 1 unit at price 760.539978, investment 3.323003 %, total balance
9257.699951,
day 14, sell 1 unit at price 768.270020, investment -0.383797 %, total balance
10025.969971,
day 22: buy 1 unit at price 762.520020, total balance 9263.449951
day 24, sell 1 unit at price 771.190002, investment 1.137017 %, total balance
10034.639953,
day 28: buy 1 unit at price 796.099976, total balance 9238.539977
day 29: buy 1 unit at price 797.070007, total balance 8441.469970
day 31, sell 1 unit at price 790.799988, investment -0.665744 %, total balance
9232.269958,
day 32: buy 1 unit at price 794.200012, total balance 8438.069946
day 33: buy 1 unit at price 796.419983, total balance 7641.649963
day 37, sell 1 unit at price 791.549988, investment -0.692539 %, total balance
8433.199951,
day 39, sell 1 unit at price 782.789978, investment -1.436670 %, total balance
9215.989929,
day 41, sell 1 unit at price 786.140015, investment -1.290772 %, total balance
10002.129944,
day 48: buy 1 unit at price 806.359985, total balance 9195.769959
day 49: buy 1 unit at price 807.880005, total balance 8387.889954
day 50, sell 1 unit at price 804.609985, investment -0.217025 %, total balance
9192.499939,
day 52: buy 1 unit at price 802.174988, total balance 8390.324951
day 53, sell 1 unit at price 805.020020, investment -0.354011 %, total balance
9195.344971,
day 56, sell 1 unit at price 835.669983, investment 4.175522 %, total balance
10031.014954,
day 67: buy 1 unit at price 809.559998, total balance 9221.454956
day 68: buy 1 unit at price 813.669983, total balance 8407.784973
day 69, sell 1 unit at price 819.239990, investment 1.195710 %, total balance
9227.024963,
day 71: buy 1 unit at price 818.979980, total balance 8408.044983
```

day 72, sell 1 unit at price 824.159973, investment 1.289219 %, total balance 9232.204956,
day 76: buy 1 unit at price 831.330017, total balance 8400.874939
day 77, sell 1 unit at price 828.640015, investment 1.179520 %, total balance 9229.514954,
day 78: buy 1 unit at price 829.280029, total balance 8400.234925
day 79, sell 1 unit at price 823.210022, investment -0.976747 %, total balance 9223.444947,
day 80: buy 1 unit at price 835.239990, total balance 8388.204957
day 81, sell 1 unit at price 830.630005, investment 0.162789 %, total balance 9218.834962,
day 83, sell 1 unit at price 827.780029, investment -0.893152 %, total balance 10046.614991,
day 88: buy 1 unit at price 845.539978, total balance 9201.075013
day 89, sell 1 unit at price 845.619995, investment 0.009463 %, total balance 10046.695008,
day 91: buy 1 unit at price 848.780029, total balance 9197.914979
day 92: buy 1 unit at price 852.119995, total balance 8345.794984
day 93: buy 1 unit at price 848.400024, total balance 7497.394960
day 96, sell 1 unit at price 817.580017, investment -3.675865 %, total balance 8314.974977,
day 97: buy 1 unit at price 814.429993, total balance 7500.544984
day 100, sell 1 unit at price 831.409973, investment -2.430411 %, total balance 8331.954957,
day 101, sell 1 unit at price 831.500000, investment -1.991988 %, total balance 9163.454957,
day 103, sell 1 unit at price 838.549988, investment 2.961580 %, total balance 10002.004945,
day 104: buy 1 unit at price 834.570007, total balance 9167.434938
day 105: buy 1 unit at price 831.409973, total balance 8336.024965
day 106, sell 1 unit at price 827.880005, investment -0.801611 %, total balance 9163.904970,
day 107: buy 1 unit at price 824.669983, total balance 8339.234987
day 108: buy 1 unit at price 824.729980, total balance 7514.505007
day 110: buy 1 unit at price 824.320007, total balance 6690.185000
day 111: buy 1 unit at price 823.559998, total balance 5866.625002
day 112, sell 1 unit at price 837.169983, investment 0.692800 %, total balance 6703.794985,
day 113, sell 1 unit at price 836.820007, investment 1.473320 %, total balance 7540.614992,
day 114, sell 1 unit at price 838.210022, investment 1.634479 %, total balance 8378.825014,
day 115, sell 1 unit at price 841.650024, investment 2.102341 %, total balance 9220.475038,
day 116, sell 1 unit at price 843.190002, investment 2.383555 %, total balance 10063.665040,
day 122: buy 1 unit at price 912.570007, total balance 9151.095033
day 123, sell 1 unit at price 916.440002, investment 0.424077 %, total balance

10067.535035,
day 125: buy 1 unit at price 931.659973, total balance 9135.875062
day 126, sell 1 unit at price 927.130005, investment -0.486225 %, total balance
10063.005067,
day 144: buy 1 unit at price 966.950012, total balance 9096.055055
day 145, sell 1 unit at price 975.599976, investment 0.894562 %, total balance
10071.655031,
day 146: buy 1 unit at price 983.679993, total balance 9087.975038
day 148: buy 1 unit at price 980.940002, total balance 8107.035036
day 149, sell 1 unit at price 983.409973, investment -0.027450 %, total balance
9090.445009,
day 151, sell 1 unit at price 942.900024, investment -3.877911 %, total balance
10033.345033,
day 155: buy 1 unit at price 939.780029, total balance 9093.565004
day 157, sell 1 unit at price 950.630005, investment 1.154523 %, total balance
10044.195009,
day 160: buy 1 unit at price 965.590027, total balance 9078.604982
day 161, sell 1 unit at price 952.270020, investment -1.379468 %, total balance
10030.875002,
day 164: buy 1 unit at price 917.789978, total balance 9113.085024
day 166, sell 1 unit at price 898.700012, investment -2.079993 %, total balance
10011.785036,
day 168: buy 1 unit at price 906.690002, total balance 9105.095034
day 171: buy 1 unit at price 930.090027, total balance 8175.005007
day 172, sell 1 unit at price 943.830017, investment 4.096220 %, total balance
9118.835024,
day 173, sell 1 unit at price 947.159973, investment 1.835300 %, total balance
10065.994997,
day 176: buy 1 unit at price 965.400024, total balance 9100.594973
day 179, sell 1 unit at price 972.919983, investment 0.778947 %, total balance
10073.514956,
day 180: buy 1 unit at price 980.340027, total balance 9093.174929
day 181, sell 1 unit at price 950.700012, investment -3.023442 %, total balance
10043.874941,
day 194: buy 1 unit at price 914.390015, total balance 9129.484926
day 195, sell 1 unit at price 922.669983, investment 0.905518 %, total balance
10052.154909,
day 197: buy 1 unit at price 926.960022, total balance 9125.194887
day 199: buy 1 unit at price 910.669983, total balance 8214.524904
day 201, sell 1 unit at price 924.690002, investment -0.244889 %, total balance
9139.214906,
day 202, sell 1 unit at price 927.000000, investment 1.793187 %, total balance
10066.214906,
day 208: buy 1 unit at price 939.330017, total balance 9126.884889
day 211, sell 1 unit at price 927.809998, investment -1.226408 %, total balance
10054.694887,
day 214: buy 1 unit at price 929.080017, total balance 9125.614870
day 216, sell 1 unit at price 935.090027, investment 0.646878 %, total balance

```
10060.704897,
day 219: buy 1 unit at price 915.000000, total balance 9145.704897
day 221: buy 1 unit at price 931.580017, total balance 8214.124880
day 222, sell 1 unit at price 932.450012, investment 1.907105 %, total balance
9146.574892,
day 223, sell 1 unit at price 928.530029, investment -0.327399 %, total balance
10075.104921,
day 224: buy 1 unit at price 920.969971, total balance 9154.134950
day 226: buy 1 unit at price 944.489990, total balance 8209.644960
day 227: buy 1 unit at price 949.500000, total balance 7260.144960
day 228, sell 1 unit at price 959.109985, investment 4.141287 %, total balance
8219.254945,
day 229, sell 1 unit at price 953.270020, investment 0.929605 %, total balance
9172.524965,
day 230: buy 1 unit at price 957.789978, total balance 8214.734987
day 231: buy 1 unit at price 951.679993, total balance 7263.054994
day 232, sell 1 unit at price 969.960022, investment 2.154821 %, total balance
8233.015016,
day 233, sell 1 unit at price 978.890015, investment 2.202992 %, total balance
9211.905031,
day 234, sell 1 unit at price 977.000000, investment 2.660559 %, total balance
10188.905031,
```

[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()
```



[ ]: