Portfolio Asset Allocation Using Reinforcement Learning

Alexandru Savoiu (savoiu)
Stanford University

Objective

Use reinforcement learning to create a portfolio with an **optimal dynamic asset allocation**, that outperforms the benchmark.
Test the performance on two portfolios:
- Long-Only Model Based Learning (no shorting)
- Delta-Neutral Model Based Learning (shorting allowed)

Data and Benchmark

ETFs (Highly liquid, low management fees)
- Equity: SPX, RTY
- Fixed Income: TLT
- Commodities: SPDR Gold Shares

Benchmark
- Static Allocation Benchmark: All ETFs, exclusively long
- Market Allocation Benchmark: S&P 500
- Oracle (chooses best performing ETF at end of each week)

Assumptions

Model the problem as Markov Decision Process (MDP)

Find the optimal policy using model based reinforcement learning

Portfolio construction
- Create "pairs" of underlyers. Within each pair of underlyers:
  -Long-Only RL: go long or stay in cash
  -Delta-Neutral RL go long, stay in cash, or go short (w/ leverage)

Model Definition
- **States**: s(t) = $\tilde{r}_S(t)$.
  The stock returns are discretized into positive and negative returns, and therefore:
  $\forall t$, $\tilde{r}_S(t) = 1$ if $r_S(t) > 0$
  $= -1$ otherwise

- **Actions**: a_t is the weight$_S$ the investor assigns to the ETF/ underlyer. Is it discretized such that:
  - Long-Only RL Portfolio: $a_t \in \{0, 0.25, 0.5, 0.75, 1\}$ (allocation between 0 and 100%)
  - Delta-Neutral RL Portfolio: $a_t \in \{-1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2.0\}$ (allocation between -100% and 200%)

- **Rewards**: R_t defined as the instant return:
  $\mathbf{R_t = a_t * r_1(t+1) + (1-a_t) * r_2(t+1)}$

  where $r_S(t+1)$ = the return of ETF, going from t to t+1

- **IsEnd** = 1_t=T, investments are done until the final investment date.

- **Discount Factor**: γ = 1 (I approached this as a CAPM Model)

Methodology

All variables are the same for both portfolios except the actions

States
- {(1,1), (1,-1), (-1,1), (-1,-1)}

Actions
- Long-Only : {(0, 0.25, 0.5, 0.75, 1)}
- Delta-Neutral: {(-1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2.0)}
- $w_1+w_2=1$ (for both long-only and delta-neutral)

Transition Probabilities
- $T(s_t, a_t, s_{t+1}) = P(s_{t+1}|s_t, a_t) = P(s_{t+1}|s_t)$
- **The investor's action has no impact on the market**

Model Used
- Model Based Learning
- **Best action defined as:**

$$\tilde{a}_t = \max_a \hat{R}(s_t, a) + \sum_{s'} T(s_t, a, s')\hat{V}_{opt}(s') = \max_a \hat{R}(s_t, a) + \sum_{s'} \mathbb{P}(s'|s_t)\hat{V}_{opt}(s') = \max_a \hat{R}(s_t, a)$$

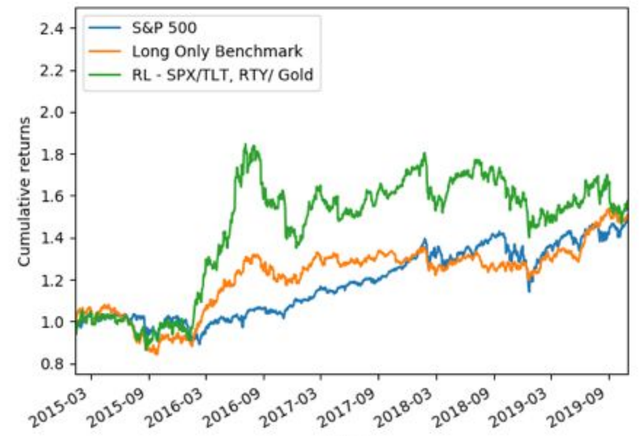
- $R_S(s_t, a)$ is an estimate of the average reward obtained when action a is taken from state s_t
- $V_{opt}(s')$ is an estimate of the expected reward of the state s', under the optimal policy
- $R(s_t, a) = \rho(s,a) / N(s,a)$
- N(s,a) stores the count of the number of times the action a was taken from the state s
- ρ(s, a) stores the cumulative sum of the previous rewards obtained every time the action a was taken from the state s

Results (overall comparison)

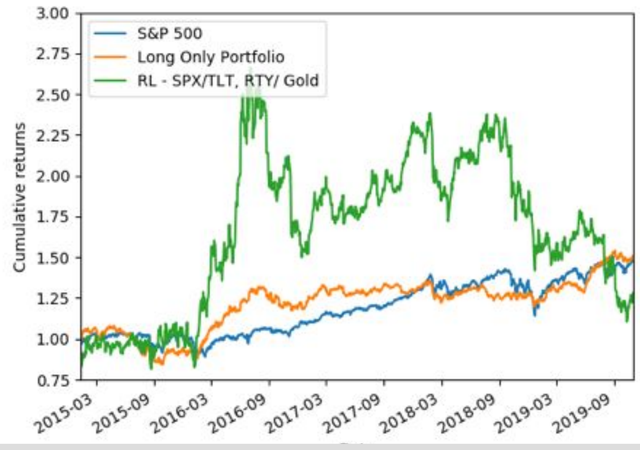
	Return	Volatility	Sharpe Ratio
Market/ S&P 500	10.48%	13.67%	0.694
Benchmark Long Only Portfolio	**10.83%**	**13.12%**	**0.749**
Oracle	35.79%	1.62%	21.33
Reinforcement Learning Portfolio – Long-Only			
RL Portfolio Long (SPY, RTY, TLT, Gold)	**12.02%**	**19.49%**	**0.565**
SPY/ TLT RL Long	7.90%	13.69%	0.504
RTY/ Gold RL Long	15.72%	31.92%	0.461
Reinforcement Learning Portfolio – Delta-Neutral			
RL Portfolio Delta (SPY, RTY, TLT, Gold)	**6.38%**	**39.71%**	**0.136**
SPY/ TLT RL Delta	8.83%	30.61%	0.256
RTY/ Gold RL Delta	3.76%	63.67%	0.044

Cumulative Returns (overall)

Long-Only Model Based Learning



Delta-Neutral Model Based Learning



Cumulative Returns (individually)

Long-Only Model Based Learning