Statistical Learning Notes

December 2019

1 Lecture 1

intro and housekeeping

2 Lecture 2

2.1 Markov Process

2.1.1 Markov Property

A given state S, is markovian iff $\mathbb{P}(S_{t+1}|S_t) = \mathbb{P}(S_{t+1}|S_1, S_2, ...S_t)$

2.1.2 State Transitions

$$\mathcal{P}_{ss'} = \mathbb{P}(S_{t+1} = s'|S_t = s)$$

$$\mathcal{P} = \begin{bmatrix} \mathcal{P}_{11} & \mathcal{P}_{12} & \dots & \mathcal{P}_{1n} \\ \vdots & \ddots & & \vdots \\ \mathcal{P}_{n1} & \mathcal{P}_{n2} & \dots & \mathcal{P}_{nn} \end{bmatrix}$$

A Markov Process (or Markov Chain) is a tuple $M = (\mathcal{S}, \mathcal{P})$

1. $\mathcal{S}$ is a finite set of states
2. $\mathcal{P}$ is a state transition probability matrix $\mathcal{P}_{ss'} = \mathbb{P}(S_{t+1} = s'|S_t = s)$.

2.2 Markov Reward Process

A Markov Reward Process is a tuple $M = (\mathcal{S}, \mathcal{P}, \mathcal{R}, \gamma)$

1. $\mathcal{S}$ is a finite set of states
2. $\mathcal{P}$ is a state transition probability matrix $\mathcal{P}_{ss'} = \mathbb{P}(S_{t+1} = s'|S_t = s)$.
3. $\mathcal{R}$ is a reward function, $\mathcal{R}_s = \mathbb{E}[R_{t+1}|S_t = s]$
4. $\gamma \in [0, 1]$ is the discount factor

2.2.1 Return

The return at a timestamp t $G_t = R_{t+1} + \gamma T_{t+2} + ... = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$
A closely related concept is the value function for a given state $v(s) = \mathbb{E}[G_t|S_t = s]$

2.2.2 Bellman Equations for MRPs

$$\begin{aligned} v(s) &= \mathbb{E}[G_t|S_t = s] \\ &= \mathbb{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + ...|S_t = s] \\ &= \mathbb{E}[R_{t+1} + \gamma(R_{t+2} + \gamma R_{t+3} + ...)|S_t = s] \\ &= \mathbb{E}[R_{t+1} + \gamma G_{t+1}|S_t = s] \\ &= \mathbb{E}[R_{t+1} + \gamma v(S_{t+1})|S_t = s] \end{aligned}$$

$$v(s) = \mathcal{R}s + \gamma \sum_{s_0 \in S} \mathcal{P}_{ss_0} v(s_0)$$

or equivalently using matrices

$$v = \mathcal{R} + \gamma \mathcal{P} v$$
$$v = (I - \gamma \mathcal{P})^{-1} \mathcal{R}$$

2.3 Markov Decision Process

A Markov Decision Process is a tuple $M = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$

1. $\mathcal{S}$ is a finite set of states
2. $\mathcal{A}$ is a finite set of actions
3. $\mathcal{P}$ is a state transition probability matrix $\mathcal{P}^a_{ss'} = \mathbb{P}(S_{t+1} = s'|S_t = s, A_t = a)$.
4. $\mathcal{R}$ is a reward function, $\mathcal{R}^a_s = \mathbb{E}[R_{t+1}|S_t = s, A_t = a]$
5. $\gamma \in [0, 1]$ is the discount factor

2.3.1 Policies

A policy π is a distribution over actions given states

$$\pi(a|s) = \mathbb{P}[A_t = a|S_t = s]$$

Note that MDP policies depend only on the current state (again history independent / memory-less).

We also observe that an MDP $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$ and a policy π together define a Markov Process $(\mathcal{S}, \mathcal{P}^\pi$, and a Markov Reward Process $(\mathcal{S}, \mathcal{P}^\pi, \mathcal{R}^\pi, \gamma$ where

$$\mathcal{P}^\pi = \sum_{a \in \mathscr{A}} \pi(a|s), \mathcal{P}^a_{ss'}$$

$$\mathcal{R}^\pi = \sum_{a \in \mathscr{A}} \pi(a|s), \mathcal{R}^a_s$$

2.3.2 Value Function

The 'state value function' $v_\pi(s)$ is the expected return starting from state s and following policy π

$$v_\pi(s) = \mathbb{E}[G_t|S_t = s]$$

The 'action value function' $q_\pi(s, a)$ is the expected return starting from state s, taking action a then following policy p.

$$q_\pi(s, a) = \mathbb{E}[G_t|S_t = s, A_t = a]$$

2.3.3 Bellman Equations for MDPs

We start by decomposing the equations for the state and action value functions

$$v_\pi(s) = \mathbb{E}_\pi[R_{t+1} + \gamma v_\pi(S_{t+1})|S_t = s]$$
$$q_\pi(s,a) = \mathbb{E}_\pi[R_{t+1} + \gamma q_\pi(S_{t+1}, A_{t+1})|S_t = s, A_t = a]$$

next we note that

$$v_\pi(s) = \sum_{a\in\mathcal{A}} \pi(a|s) q_\pi(s,a)$$

and

$$q_\pi(s,a) = \mathcal{R}_s^a + \gamma \sum_{s'\in\mathcal{S}} \mathcal{P}_{ss'}^a v_\pi(s')$$

Substitution one in the other both ways yields

$$v_\pi(s) = \sum_{a\in\mathcal{A}} \pi(a|s)(\mathcal{R} + \gamma \sum_{s'\in\mathcal{S}} \mathcal{P}_{ss'}^a v_\pi(s'))$$
$$q_\pi(s,a) = \mathcal{R}_s^a + \gamma \sum_{s'\in\mathcal{S}} \mathcal{P}_{ss'}^a (\sum_{a'\in\mathcal{A}} \pi(a'|s') q_\pi(s',a'))$$

We can once again express these functions concisely in matrix form:

$$v_\pi = \mathcal{R}^\pi + \gamma \mathcal{P}^\pi v_\pi$$

2.3.4 Optimal things

We define the optimal state value function $v_*(s)$ as the maximum value function over all possible policies.

$$v_*(s) = \max_\pi v_\pi(s)$$

Similarly, we define the optimal action-value function $q_*(s,a)$ is the maximum action-value function over all policies

$$q_*(s,a) = \max_\pi q_\pi(s,a)$$

The optimal policy is generally what we are searching for in an MDP, we consider an MDP 'solved' when we know the optimal value function.
We define a partial ordering on all policies π s.t.

$$\pi_1 \geq \pi_2 \; if \; v_{\pi_1}(s) \geq v_{\pi_2}(s) \forall s \in S$$

Theorem: For any markov decision process

- There exists an optimal policy π_* s.t. $\pi_* \geq \pi$ for any valid policy π
- all optimal policies achieve the same value function $v_{\pi_{*1}}(s) = v_{\pi_{*2}}(s) \forall s \in S$
- all optimal policies achieve the same action-value function $q_{\pi_{*1}}(s|a) = q_{\pi_{*2}}(s|a) \forall s \in S$

Given an optimal action value function or value function, we can easily find an optimal policy:

$$\pi_*(a|s) = \begin{cases} 1 & a = \underset{a\in A}{argmax} \; q_*(s,a) \\ 0 & otherwise \end{cases}$$

Lecture 3, Planning By Dynamic Programming

2.4 Iterative Policy Evaluation

Goal: evaluate a given plicy π for a MDP $M =< \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma >$. We iterative apply the Bellman expectation backup $v_1 \to v_2 \to ... \to v_p i$.

$$v_{k+1}(s) = \sum_{a \in \mathcal{A}} \pi(a|s)(\mathcal{R}_s^a + \gamma \sum_{s \ \in \mathcal{S}} \mathcal{P}_{ss'}^a v_k(s'))$$

$$\boldsymbol{v}^{k+1} = \boldsymbol{\mathcal{R}}^\pi + \gamma \boldsymbol{\mathcal{P}}^\pi \boldsymbol{v}^v$$

The convergence of this iterative algorithm is based on a fixed point argument. Nameley that the point $\boldsymbol{v}_{pi}$ is a fixed point up this update equation, (which should be clear from the bellman equations), and moreover that this update function is contraction mapping. Therefore, convergence to $\boldsymbol{v}_\pi$ is guaranteed for *any* starting point v_0.

2.5 Iterative Policy Improvement

Now that we have a way to evaluate a given policy, we would like to be able to improve it as well. we do this in a two step process
Given a policy π

1. evaluate the policy π_k via iterative policy evaluation
   $v_1 \to v_2 \to ... \to v_p i$.
2. improve the policy by acting greedily with respect to v_π

$$\pi_{k+1}(a|s) = arg \max_{a \in \mathcal{A}} q_\pi(s, a)$$

3. iterate

Again, this procedure will always converge to the optimal policy π_*, with the proof for this again being based on fixed point / contraction mapping arguments.

2.6 Value Iteration

It turns out we can actually shortcut the policy iteration portion of our iterative improvement algorithm. This should intuitively make sense as given any value function, the optimal policy is determined using a greedy strategy.

$$v_{k+1}(s) = \max_{a \in \mathcal{A}} (\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v_k(s'))$$

$$\boldsymbol{v}_{k+1}(s) = \max_{a \in \mathcal{A}} \boldsymbol{\mathcal{R}^a} + \gamma \boldsymbol{\mathcal{P}^a} \boldsymbol{v}_k$$

3 Why This Works: Contraction Mappings and Fixed Points

Define the *Bellman Expectation Backup* operator T^π as :

$$T^\pi(v) = \mathcal{R} + \gamma \mathcal{P}^\pi v$$

This function is a γ contraction i.e.

$$\begin{aligned}||T^{\pi}(u) - T^{\pi}(u)||_{\infty} &= ||\mathcal{R} + \gamma\mathcal{P}^{\pi}u - (\mathcal{R} + \gamma\mathcal{P}^{\pi}v)||_{\infty} \\ &= ||\gamma\mathcal{P}^{\pi}(u - v)||_{\infty} \\ &\leq \gamma||\mathcal{P}^{\pi}||_{\infty}||(u - v)||_{\infty} \\ &\leq \gamma||(u - v)||_{\infty}\end{aligned}$$

Similarly, if we define

$$T^{*}(v) = \max_{a \in \mathcal{A}} \mathcal{R}^{a} + \gamma\mathcal{P}^{a}v$$

we have

$$\begin{aligned}||T^{*}(u) - T^{*}(u)||_{\infty} &= ||\max_{a \in \mathcal{A}}(\mathcal{R}^{a} + \gamma\mathcal{P}^{a}v) - (\max_{a \in \mathcal{A}}(\mathcal{R}^{a} + \gamma\mathcal{P}^{a}v))||_{\infty} \\ &\leq \gamma||(u - v)||_{\infty}\end{aligned}$$

Thus convergence is shown.

4 Risk Aversion through Utility Theory

4.1 Definitions

- Utility of consumption $U(x)$
  - x represents the uncertain outcome being consumed
  - $U(\cdot)$ is a concave function, therefore $\mathbb{E}[U(x)] \leq U(\mathbb{E}[x])$
- **Certainty-Equivalent Value** $x_{CE} = U^{-1}(\mathbb{E}(U(x))$
- **Absolute Risk-Premium** $\pi_A = \mathbb{E}[x] - x_{CE}$
- **Relative Risk-Premium** $\pi_R = \frac{\pi_A}{\mathbb{E}[x]} = \frac{\mathbb{E}[x] - x_{CE}}{\mathbb{E}[x]} = 1 - \frac{x_{CE}}{\mathbb{E}[x]}$

4.2 Calculating Risk-Premium

From here on we will call $\mathbb{E}[x] = \bar{x}$ and $Var(x) = \sigma_x^2$
First we take the second order taylor expansion of $U(x)$ around $\bar{x}$:

$$U(x) \approx U(\bar{x}) + U'(\bar{x})(x - \bar{x}) + \frac{1}{2}U''(\bar{x})(x - \bar{x})^2$$

next we take the first order taylor expansion of $U(x_{CE})$ around $\bar{x}$:

$$U(x_{CE}) \approx U(\bar{x}) + U'(\bar{x})(x_{CE} - \bar{x})$$

Taking the expectation of $U(x)$ we get

$$\mathbb{E}[U(x)] \approx U(\bar{x}) + \frac{1}{2}U''(\bar{x})\sigma_x^2$$

Noting that $\mathbb{E}[U(x)] = U(x_{CE})$ we have

$$U'(\bar{x}(x_{CE} - x) \approx \frac{1}{2}U''(\bar{x}) \cdot \sigma_x^2$$

From this, we can get expressions for the Absolute and Relative Risk Aversion

$$\pi_a = \bar{x} - x_{CE} \approx \frac{1}{2} \cdot \frac{U''(\bar{x})}{U'(\bar{x}} \cdot \sigma_x^2$$

$$\pi_r = \frac{\pi_a}{\bar{x}} \approx \frac{1}{2} \cdot \frac{U''(\bar{x}) \cdot \bar{x}}{U'(\bar{x}} \cdot \frac{\sigma_x^2}{\bar{x}^2} = 1\frac{1}{2} \cdot \frac{U''(\bar{x}) \cdot \bar{x}}{U'(\bar{x}} \cdot \sigma_{\frac{x}{\bar{x}}}^2$$

Define:

1. **Absolute Risk-Aversion** $A(\bar{x}) = -\frac{U''(\bar{x})}{U'(\bar{x})}$
2. **Relative Risk-Aversion** $R(\bar{x}) = -\frac{U''(\bar{x}) \cdot \bar{x}}{U'(\bar{x})}$

4.3 CARA and Applications

4.3.1 CARA definition

Allow $U(x) = \frac{-e^{-ax}}{a}$ for $a \neq 0$

$$A(x) = \frac{-U''(x)}{U'(x)} = a$$

a is called the coefficient of Constant Absolute Risk Aversion (CARA) if we allow the randome outcome $x \sim \mathcal{N}(mu, \sigma^2)$, then

$$\mathbb{E}[U(x)] = \begin{cases} \frac{-e^{-a\mu + \frac{a^2\sigma^2}{2}}}{a} & a \neq 0 \\ \mu & a = 0 \end{cases}$$

$$x_{CE} = \mu - \frac{a\sigma^2}{2}$$

Therefore, $\pi_a = \frac{a\sigma^2}{2}$

4.3.2 CARA applied to portfolio allocation

consider the following scenario

- We are given \$1 to invest and hold for a horizon of 1 year
- Investment choices are 1 risky asset and 1 riskless asset
    1. riskless asset annual return $\sim r$
    2. risky asset annual return $\sim \mathcal{N}(\mu, \sigma^2)$
- we want to determine the optimal (unconstrained) π to allocate to risky asset $(1 - \pi)$ is allocated to riskless asset to maximize utility of wealth in 1 year

We note that portfolio wealth $\mathcal{N}(1 + r + \pi(\mu - r), \pi^2\sigma^2$ so we optimize this, i.e. differentiate and set equal to zero yielding

$$\pi^* = \frac{\mu - r}{a\sigma^2}$$

4.4 CRRA and Applications

4.4.1 CRRA definition

Allow $U(x) = \frac{x^{1-\gamma}}{1-\gamma}$ for $\gamma \neq 1$

Relative Risk Aversion: $R(x) = \gamma$, γ is the *Coefficient of Constant Relative Risk Aversion*, note for $\gamma = 1$, $U(x) = log(x)$

If the randome outcome x is lognormal, $log(x) \sim \mathcal{N}(\mu, \sigma^2)$

$$\mathbb{E}[U(x)] = \begin{cases} \frac{e^{\mu(1-\gamma)} + \frac{\sigma^2}{2}(1-\gamma)^2}{1-\gamma} & gamma \neq 1 \\ \mu & \gamma = 1 \end{cases}$$

Relative Risk Premium, $\pi_R = s - \frac{x_{CE}}{\bar{x}} = 1 - e^{-\frac{\sigma^2\gamma}{2}}$

4.4.2 CRRA application, Portfolio Construction (Merton 1969)

Problem Definition, Merton's 1969 Portfolio Problem

- 1 risky asset, 1 riskless asset
- Riskless asset $dR_y = r \cdot R_t \cdot dt$
- Risky asset $dR_y = \mu S_t \cdot dt + \sigma \cdot S_t \cdot dz_t$, (Geometric Brownian, more on this later)
- Given \$1 to invest, continuous rebalancing
- Determine π, fraction of W_t to allocate to risky asset to maximize expected utility of wealth $W = W_1$

The process for wealth with this construction is:

$$dW_t = (r + \pi(\mu - r)) \cdot W_t \cdot dt + \pi \cdot \sigma \cdot W_t \cdot dz_t$$

Solve with CRRA Utility $U(W) = \frac{W^{1-\gamma}}{1-\gamma}$, $\gamma \in (0, 1)$

Applying Ito's Lemma on $log(W_t)$ gives:

$$log(W_t) = \int_0^t (r + \pi(\mu - r) - \frac{\pi^2\sigma^2}{2}) \cdot du + \int_0^t \pi \cdot \sigma \cdot dz_u$$
$$\rightarrow log(W) \sim \mathcal{N}(r + \pi(\mu - r) - \frac{\pi^2\sigma^2}{2}, \pi^2\sigma^2)$$

From the previous section, we need to maximize :

$$r + \pi(\mu - r) - \frac{\pi^2\sigma^2\gamma}{2}$$

therefore

$$\pi^* = \frac{\mu - r}{\gamma\sigma^2}$$

5 Stochastic Calc Primer

Taking a step back to do some math to understand what's happening...

5.1 Intro

Let's begin by looking at a basic random process. Consider a series of coin tosses. At each toss, a heads means a win of \$1, a tails means a loss of \$1. Let R_i be the outcome of the ith toss.
Clearly, R_i is a random variable:

$$\mathbb{E}[R_i] = 0,\ \mathbb{E}[R_i^2] = 1,\ \mathbb{E}[R_i R_j] = 0$$

Now, let $S_i = \sum_{j=1}^{i} R_j$, S_j is an example of a random walk.

$\mathbb{E}[S_i] = 0$, $\mathbb{E}[S_i^2] = \mathbb{E}[R_1^2 + 2R_i R_j + ...] = i$ importantly, however, $\mathbb{E}[S_i|S_1...S_j] = \mathbb{E}[S_i|S_j] = S_j$.

This process exhibits both the markov property and the martingale property.

5.2 Brownian Motion

6 HJB Equation and Merton's Portfolio Problem

References

- http://web.stanford.edu/class/cme241