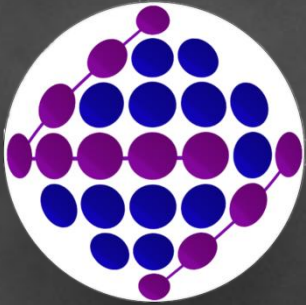


Reinforcement Learning



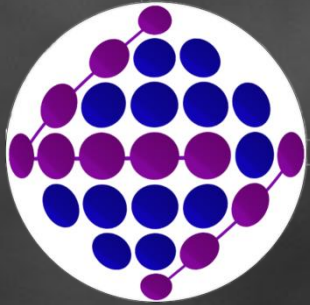
فصل نهم



شرکت فروش مصنوعي رسا

مدرس : فرید فاشمی نژاد

Policy Gradients



جلسہ پنجا لقم



شرکت فروش مصنوعي رسا

مدرس : فرید فاشمی نژاد

- سیاست‌گرایان ما
- مدل قبلی خیلی خوب نبود
- شاید چون به تاریخچه اقدامات توجه نشده
- و فقط اقدام قبلی را در نظر گرفتیم
- به این مسئله assignment of credit می‌گویند.

- ما این مسئله را با اعمال discount rate، این نرخ را بین 0.95 تا 0.99 در نظر میگیرند و محاسبه میکنند.
- امتیاز:

$$R_{t=0} + R_{t=1}D + R_{t=2}D^2 + \dots + R_{t=n}D^n$$

• مثال :

ستون بالا می رود ستون بالا می رود ستون پایین می رود



R:1
-100

L:0
+10

R:1
+10

اقدام

Sum DR = -70.75

$$= 10 + (0.95) * 10 + (0.95)^2 * (-100)$$