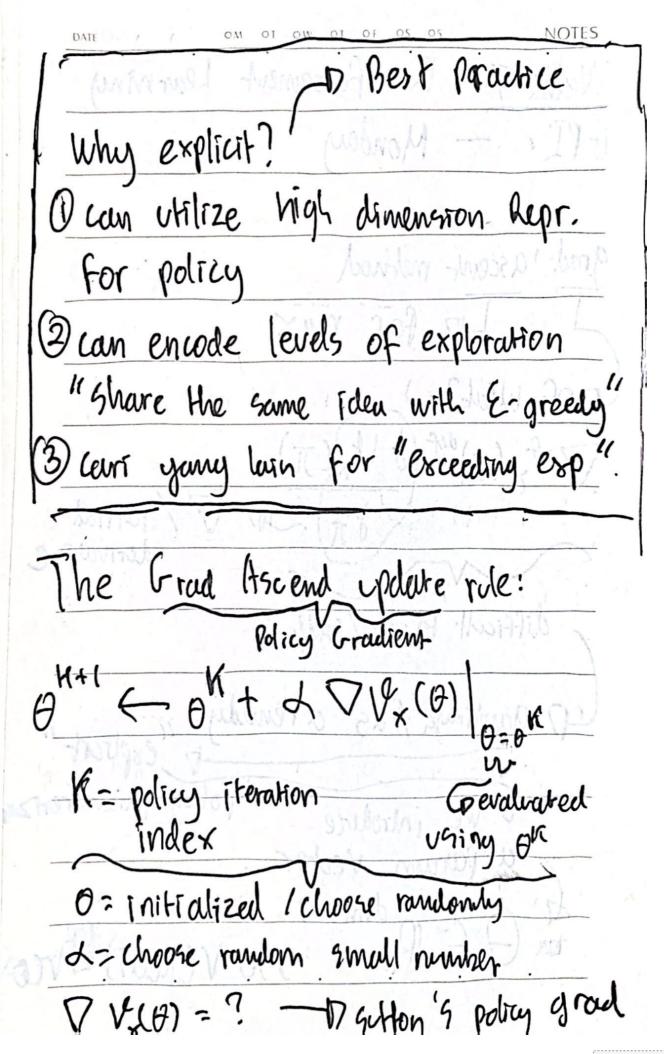
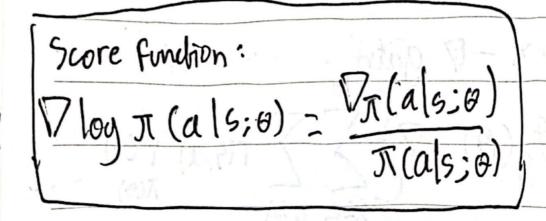
Week	7 -	Kem f	orcemu	A Leav	rminy
GPI	_	Monda	y	halles	r Jodel
.193	NOT	restant.	HIV	selle:	100
grad o	scent	method			707
(195	Lp	for	MUX	3/10/14	s mai (C)
pof u	that?	v do	5.30	041	repaire "
∇v_{x}	(J) de	() Vx	(JT)	1 parage	. 7000 (E
	~	187	1~10	√/	fartial derivative
, diff	scult t	o use lia	aldly		
-17 ga	Minya	/ as	u reml		splient"
6	We	introlluce		potrcy 1	parameter 2
1 a	Para	introlluce m veo	for	AT 3, 122 U	
4 (121	D dim (6	2)	100	al) def
MT C		1	190	V (ICC	71) ~ VCE



 $V_{q}(\theta) = \nabla \left\{ \sum_{r(s,a)} p^{*}(s) \right\}$... R(alsit)... by def of gain y = \(\int \text{T (S,a)} \) \(\text{Tp* (S)} \) \(\text{T} \) \("Score function". = \(\sigma\) \(\frac{1}{27}\) \(\frac{1} = 2 2 p*(5). T(a/5)-1(4,a) { Ploy p* + 7 log T? Lp Gampling-freendly but in RL, P* 13 still unknown to the agent.





= \(\sum_{5} \omega \pm^*(5) \omega \tals) \quad \(\begin{align*} \quad \text{S,a) \omega \text{log T (als)} \\ \end{als}\)

= E [9] (SIA) V log T (A 15:18)]

5~p* Au~7(.(5) Lipexceeding exp

MALINIA N.

p* is unknown but appears under

E, hence in prouble we simply

Sample forom Pr (6)

Sperest 1 at tytmix

run unbiased approx of 712 (x)

but,	ar t	< fmix,	5	- pt ≠ pts	-
herve	biased	of prox.	08	Duck)	

(D.BCV: Reduce Var w/o introducing

hias error

baseline

7 Vg(0) = E[{97 (5,A) - V; (5) Y7loy -...

TI (A15:0)]

Factorwage function

= 2 T (5,A)



8, (4, A, 5') V lay π (A(5;0))



Weeh 7 -	Reinforce m	rent Lea	rning
Wednesday.	Will Gore	brown	Lomital
Marine Service	14 rog Japans	P. Hora	11

Polity param for Discrete Adjon (focus: finite MDP)

Using: (Fibbs/Boltzman/questhaux categorical dist softmax Parameterization

J(als; θ): exp(θ^T β(s,a)) +(s,a)6(sxA) Σ exp(θ^T β(s,a))

The focus of the

(0: 1/A) 2017

Discounted Reward Policy	Galadient
Recall: 27 100 4 ymbol Fe	
18 (JT) = Por To Vg (J	T)= (7 7
	(1914 550)
	1727
PX = lim \(\S.Px)+\P	(500 Drew)
+ mix-720 +	T -CKC (100)
Lpis not as proper ds	stribution
beruse $\geq P_{\pi}^{8}(516) \neq 1$	
4:0	
VVox (Tr, 50) = [(5/5/5) >	$= Q_{x}^{\pi}(S_{x}a)$
·· Vlayπ(Al); ¿θ)	Ä
5	1711 60
not sampling friendly,	not a distributing

times with the (1-2)
$=\frac{1}{(1-8)}\sum_{l}(1-8)P_{\pi}^{8}(5 50)\sum_{u}$
Lo hecomes a geometric
Distribution
Distribution Distribution to be made it back to equalify.
Tg = arg max (by (JT, 40) TET
Ty=argmax E[V; (T,50)] The argmax E[V; (T,50)] The argmax E[V; (T,50)]



	All the first party and			(6.1.
13ms (136)	will a	onally by		
	rollion	rieW.		
e genelite	at No.	divio	adamir.	O
			1	
14	101	(11) B	y your	P10 =
March !			THE	

