PAGE NO .

DATE: / /

OPTIMISATION ALGORITHMS

	Video 1: Mini batch gordient descent
*	If m= 500 so million then training such a huge data set is difficult and applying graduent descent makes the algorithm slower
	data set is difficult and appleans graduent
	descent makes the adopathm flower
	V. Company of the Com
*	Solution: Splitting me to mini batches of size
	1000, and annume this to X24
*	X = X(1) X(2) X (1000) X 1001
	(nr,m)
5 -	x(1) (nx,1000) x(2) (n,1000)
101.	
1 1	$y = (y^{(1)}, y^{(2)}, y^{(3)}, y^{(100c)}, y^{(100c)}, y^{(200c)}, y^{(200c$
((, m)
	(1) (1, 1000)
	1 (1,100)

minibatch t: x 2+3 y 2+3

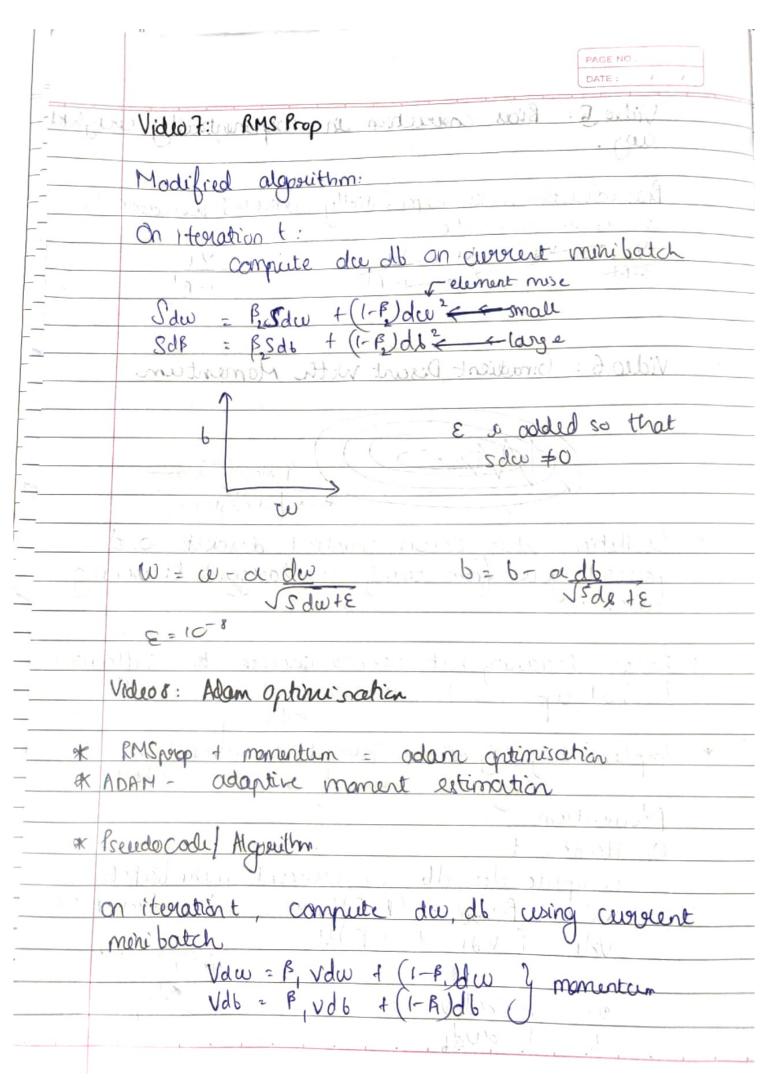
doute sets. Thus, we can calculate it for smillion examples at some time asing vectors when

	DATE: / /
	Algorithm: 219H7.19001A 110IIIA211111111111111111111111111111111
	Stepeat 2 Him buton graduant hurst
1-	forward prop on X Et 5
(g) S	i i i i i i i i i i i i i i i i i i i
	ACD - GCO Z CO
1 1 3 T V	compute cost J 25 1000 2 2 (g', g') + 2 5/16/1
	backpage to compute gradient w. A. t J (th)
	ber = ber - ade (1)
	3
	July 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

PAGE NO .
DATE: / /

Cost		radiat o	la scount						
		4			Min	n, pa	tch g	radio	ent d
cost	1.	7 7.	77.	n : 1	+	, Y	× 113	41	[13
cost	1	96113		11	M	1	× 123	,	223
					nuy	(,		, , ,	
``				cost		My		(
16.2	A vera	Cold S	Wes	tialloid	1500	YÜ!	Ming	min	113
	#i	teration		7 3(] -1	1.	mind	batch	# t	
		- 1	1						
Choc	sing y	our min	vi bate	h size	2:	Bro	le Vi	.2)	
4	0			\	31	200 (0	ng		
mi	ni batan	size:	m → (botch &	ractie	nt o	lescin	t	
						0000		00.00	U
		between	() c @ →	mini	bate	h	growl	ient	dex
1:4	200	, 100 1121					John Jr.	FAIC	
117		too noisy				9	nair		
	1 3 7 1 1 (ose spee	dup	no 6_	المد	-(3)	Fie		
1.,	1. 1. 0.0	vanit su	ach m	un co	st	1/0		, , , ,	
7					Sulli		· la		
				- 2					
min't	ratah	gradient	ander	entiebr	07810	bral) : 3	lidge 1	j
0 81	254091	Leaunin	a:	U			WE.		
-	Location	leavening isochion	Boliva	ntaso					
> V	m Kr	pologoress training	culto	Mark	(1)0	tilac	1	racee	H
0.0	500	training	Cot	, 40,50	www.		- 60	o Carl	on
· do	pant o	STOREHLI	Cania	eme C	Tiere!	lat	7.1	No.12	
	1) cm	salty	COLUMN C	00	م ما م م	all	المد	In Z	ma
gle	g (on	NOU C	jour o	-VC 9(LOUI C		al)		
		D I I	ii a		<u>.</u>		1	-	
		po 3		177.10	DU.	λ,	011	10	

	PAGE NO DATE: / /
- 1,	guidebres four bloosing mini baten \$1220 0261
-	(i) of teraining set size is less than 2000
-	we batch gradient descent.
	(3) 14 1001 1) 14 4001110) 01 2
-	(iii) Make cure it file in EN (GIPU momony
-	
-	Video 3: Exponentially Weighted Averages
	V = BV+1 + (1-\$)6+
	temporature sois which many property
-	
	1 could tribute that a the contract of many
	hout tollow days tota & O) and or
-	I t comp with both - 3 report to
-	Exponentially weighted averages are useful for
	optimizing graduent descent algorithm. It gives
	different & band on B. This enduces oscillation
	in gradient descent and makes smooth path
	towards minima
	Video 4: Understanding Exponentially weighted
	Averages
	and such as the state of the
1 1	VA) = BV++ + (1-B) O+ NOR
	to realize the section
1	overage à supresented over 1 entries.
	action of the contract of the
	best beta is vitueen 0.9 & 0.9P
	Va = 0



	PAGE NO.						
	DATE: / /						
· COP	Sdw= B, sdw + (-B2)dw } RMS prop						
	S do = B2 Sdo + (1-B)d62						
4.	men with the doll is you this and processing						
	The street of th						
N.	MADOS COMPONIO COMO .						
	Convicted						
	VAND = Vde de						
	(1-Bt) (1-B)						
	A Company of the Comp						
9	convicted = Sde = Sde						
	$\frac{1}{1-\beta_2} = \frac{1}{1-\beta_2}$						
	W: W- or Vdw b= b- or Vdb Sdu + E Sdu + E						
	W: W- or Vdw b= b- or Vdb Sdu + E Sdu + E						
,							
111.12	to poor the least of a for principal of ay a la						
	the lange promise on a second						
	- pullou unit						
Ho	therbroughts good to unposed off : 01 orph						
1201.	a - toute tuned a party level language						
	B, 2 0.9 (da)						
	B, = 0.99 (dw)						
	Rosman digit of 1 star house 501 = 3						
(f.	N. 1						
Vic	les 9: Leavining Rate Decay						
	U U						
W	vile complementing muni batch graduent duscur u steps will Obe nosy and it wint converge						
you	a steps will Obe many and it wint concerns						
at	minimum						
-							

PAGE NO .

DATE: / /

	but as you suduce out with time at the beginning leaving rate would be fast but then as a keeps decreasing ha will help ascillate in highter sugion arow
	leaving rate would be fast but then as a keeps
	decreasing na will help oscillate in highter ougin about
	minimum
	SERVER CENTRAL OFFICE OF SER.
	Q 2 1 do
	1+ decay rate * epochnumber
	[0
	Other methods: hyperparameter
	3 K C
	a = 0.95 epoch number x do
	a = K X a
	Sepoch number
	3 7 3 3 3 4 3 5 7 7 7 7
	charges to learning rate can in made discretely- declase after some no of epochs otherwise
	Manually.
	Video 10: The problem of local optima
	The normal local optima is not likely to appear
	in a deep NN
. 4	The Graph on a second
	Patrau- eligion of slaw harring 3
	Togother Joseph
	Weller as I consider Rober Brown
	Video 4: Leasung Robe Decery
	who harmony what many printegraphy willy
	to be the term of the second o

	Video S: Bias connection in exponentially weighted
	aws.
	of the side by the his
	Bias correction makes exponentially weighted averages & most
	and moon accurate.
-	Vt = Vt Vt Vt Vt Vt - Vt
	11 - 12 - 13 - 13 - 13 - 13 - 13 - 13 -
	2 11 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	Video 6: Gradient Descent With Momentum
	VICTO 6. Oraquere Descrit volte regradue
	(AAA)
	(Slower learning
	(Juster learning
*	Orillations eleca double amplicat descent and
	convert view from maine a lagrage black ourse
	Oscillations slow down gradient descent and powert you from cessing a larger leavening rate.
	1000.
æ	Lazzer leaguer with come ld course to illation
4	to shoot up. without with a mild it outil
	Viduo 5: Aven Ephine: rakin 1 ap toute of
*	Touch and board Consider the Hill was a 1919
1	Implementing Gradient descent + momentum
	Manny Aug 20 Man Aug 20 Man 4
	Momentian:
	On iteration t:
	compute du db on current minibatch
-	Vdw=BVdw+ (1-B)dw Vois Bret + (-B)E
	Vdb = BVdb + (1-B)db
	Mynuparametrs
1	cu = (e-adeu?) 1 1/2 ? the gods of
	b = b-audb