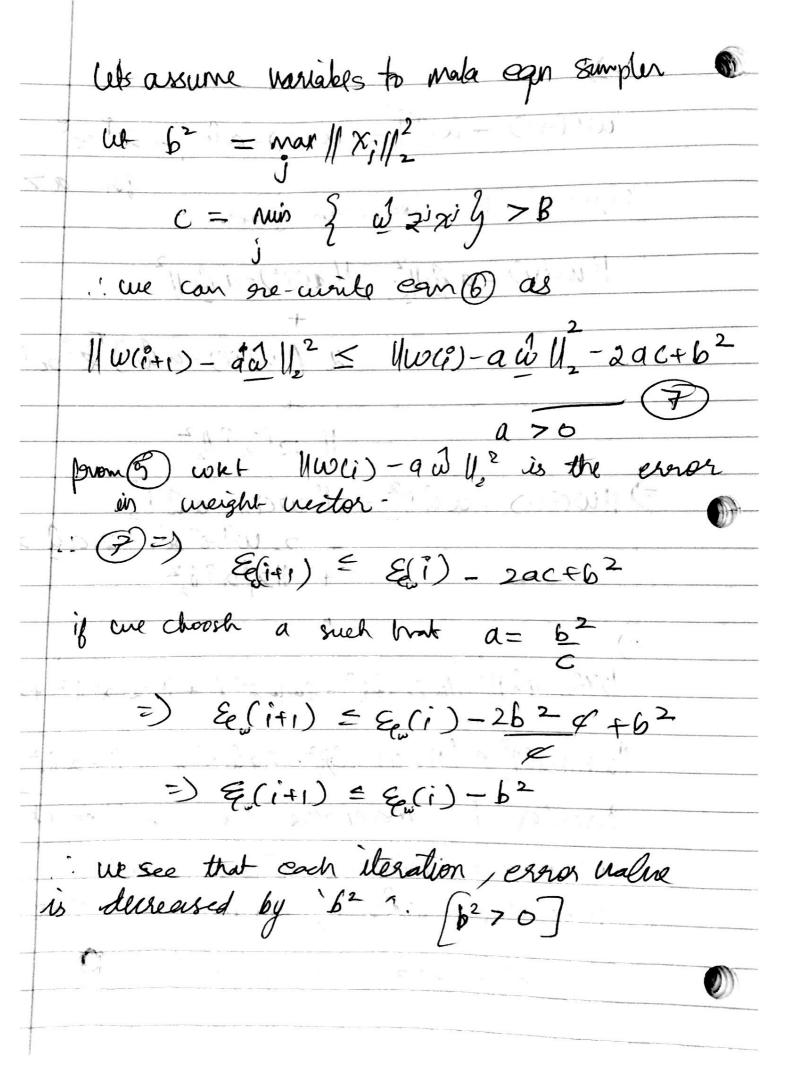
	Marginal Perceptron Lonvergence Proof
	THE CONCE COURS LICENS
	In this proof, the following assumptions are made
-	- We can short the Marinel Percepti
	* Fined inisement, q(i) = n = constant >0
	for ease we assume n = 1 [generality
	* Fined inisement, $q(i) = \eta = constant > 0$ for ease we assume $\eta = 1$ [generality []
	* All the data points are considered to be linearly
	seperable = = +(300 = (1+1)00
	when weight & B. He
	& sequential Wadient Descent technique
	inviseder (ration house) of in the
	Marchen a diff Marchine City
	* w(o) is varleitary.
	thought is also a solution
1	$w(i+1) = w(i) + z_i x_i q$ where $w(i) z_i x_i < B$
	when 2:x; represents the reflected data points.
	the current cuchinally ordered was
-	and this egres is iterated over many epochs.
	from () as n=1, we can rewrite themas
	$W(l+1) = W(l) + Z_i \chi_i \qquad -3$
-	Ut us make another assumption
-	= 12 where i=91,2c be the subset of
-	J'sé where i=91,2c be the subset of braining data points that are misclarified

at each epoch's iteration. in this proof the following assurations as me . We can start the Marginal Perceptron transporter mas (i) o more than the or ease we assume a= w(o) = arkitary value when with the wife of the will the selection with the selection when wist = 12? = B ti & segmential Wadsent Descent technique of in is a (weigh victor) solutions i.e co zaziodzaBi ota 1 then a w is also a solution a lie a with SBU = Will [Kazo] We know that the weight wester w has some error value prot beeps decreasing in each iteration. Let that error value be defined by Ew(i) = 1 Win - aw 112 from equiations 3 Ray, we can $W(i+1) - a \vec{\omega} = W(i) + z^{2} = a \vec{\omega}$

(w(8+1) - a) = (w(i)-a) /+ 2 2 11 w(i+1) - a w/ = 11 w(i)-a w 11 2 (wer)-aw) z'n' $||z'x'||_{2}$ $= ||w(i+1) - aw||_{2}^{2} = ||w(i) - aw||_{2}^{2}$ $+ 2w^{T}z^{2}z^{2} - 2awz^{2}z^{2}$ 11 2 x 1 12 + 1/2/21/2 |\w(i+1)-a\w)|2=(|\w(i)-a\w)2-2a\wzix+11=ixi112+2w\zixi $\|w(i_1) - a \hat{w}\|_{2}^{2} = (\|w(i) - a \hat{w}\|_{2}^{2} - 2a \hat{w} \hat{z}^{i} \hat{z}^{i} + \|\hat{z}^{i} \hat{z}^{i}\|_{2}^{2}) + 6$ Enoder to Maximize the equation at a we need to minimize atzizi and maximize 1/2/2012 value such that ŵ z'x'>B and 12°x'1, > ₽

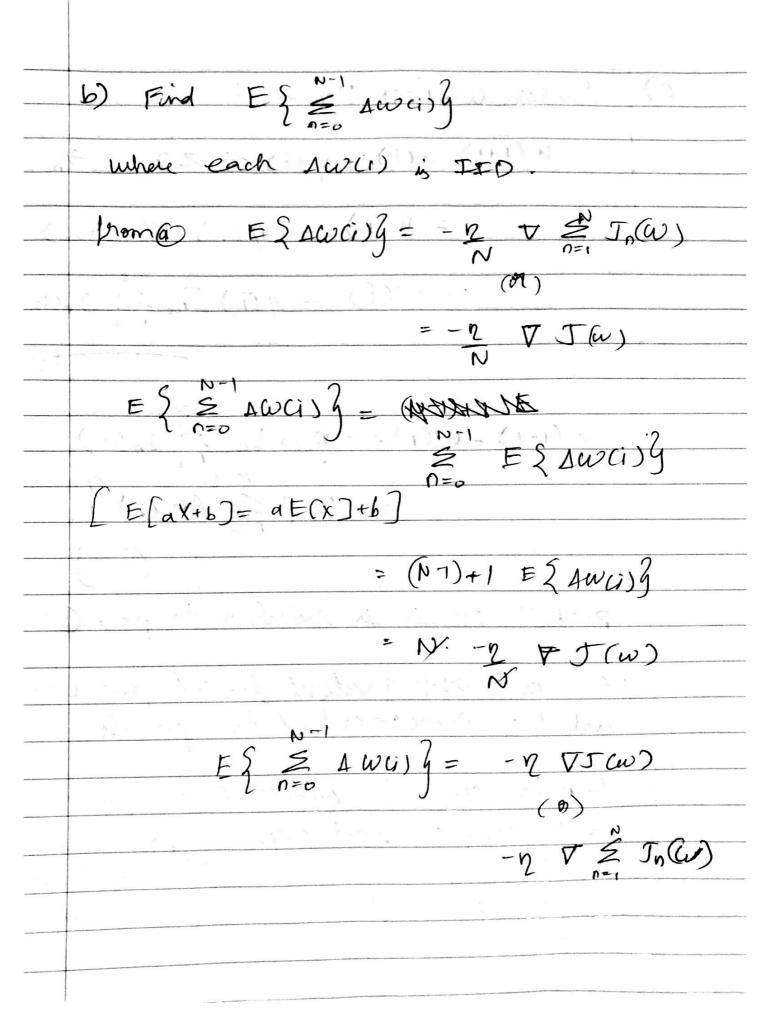


i ue can curille as
0 = 82 9. (is) = 8. (i)-1,2
0 5 2 (1) 5 En (1) -62 -8
 at some & iteration (i), Ew (i) = 62
: car @ keromes implanible
Mester the stragithmi's iteration stone of
(1-1) h i les abon and the Massis Donding
Therfor the algorithm's iteration stops out (11)th i teration, and the corresponding weight victor is the connerged weight victor.

$$P(n) = \frac{1}{(N+1)+1} = \frac{1}{N}$$

$$P(n) = \frac{1}{(N+1)+1} = \frac{1}{N}$$

$$P(n) = \frac{1}{N} = \frac$$



Batch Condient Descent: $w(i+1)=w(i)+y(i) \leq z_n x_n + v_n$ = w(i) - 2(i) To J(w) = w(i) - q(i) Tw \subseteq Jo(w) w(i+1) - w(i) = - y(i) Tw = In(w) = E \ \frac{5}{r=0} Awij \gamma part @ result is similar to part 6 1.9 in butch gradient desient we consider all the misclassified data points but in Stochastic gradient, we consider single sample and repeat for N'samples. It is computationally stable because it can be used for a set inch more minimal