

Stochastic Gradient Descent (500) 2
Approximate the boss gradieix using justs one donta instance
before:  dL(w) = 1 = dl(fw(an), yn)  dw = N = dw
With SDG
dL(w) a dl(fw(ne), yr) dw dw
$\Rightarrow \omega^{(\tau+1)} = \omega^{(\tau)} - \lambda  dl(f\omega(n\epsilon), y\epsilon)$
where $T = current$ iteration $(n_T, y_+) = (n_n, y_n) & current$ iteration

## Mini-batch

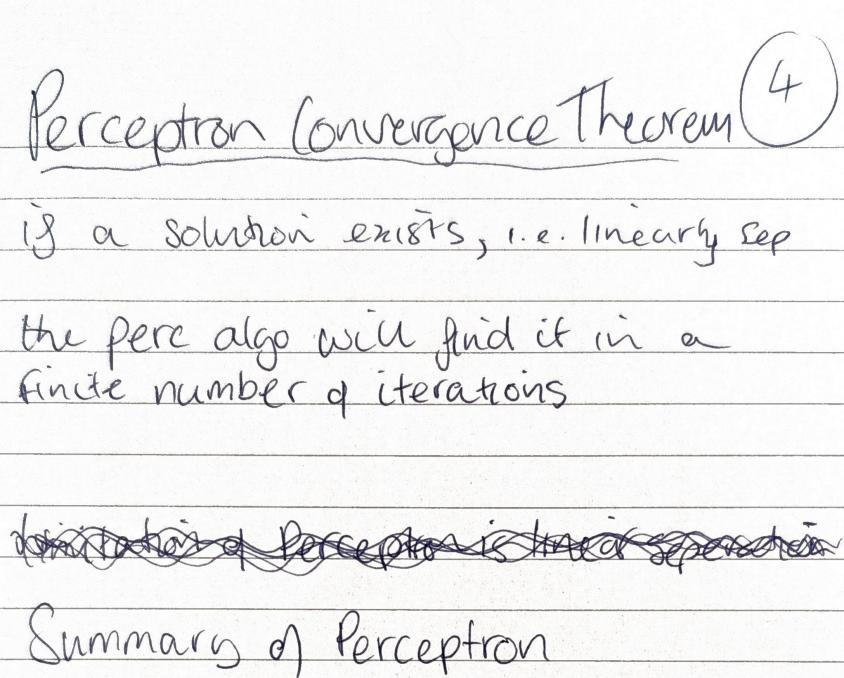
Ests sonewhere in between carred Mini-batching

MB updates w/ approximate descent of B data instances

{(x<sub>b</sub>, y<sub>b</sub>), y<sub>b</sub>, 1 ½ b ½ β g<sub>T</sub>, ω(ere β λ N) \* ω(τ+1) = ω<sup>(τ)</sup> - λ g Z dl(fw(x<sub>b</sub>), y<sub>b</sub>)

dω

under the model using the loss gradient from this batch



- Dit is a single layer NN
- (2) Only wortes for binary classification
- 3 Perception will fail for non-lineary Seperable classes
- (4) Optimization of model paras based on gradient Descent