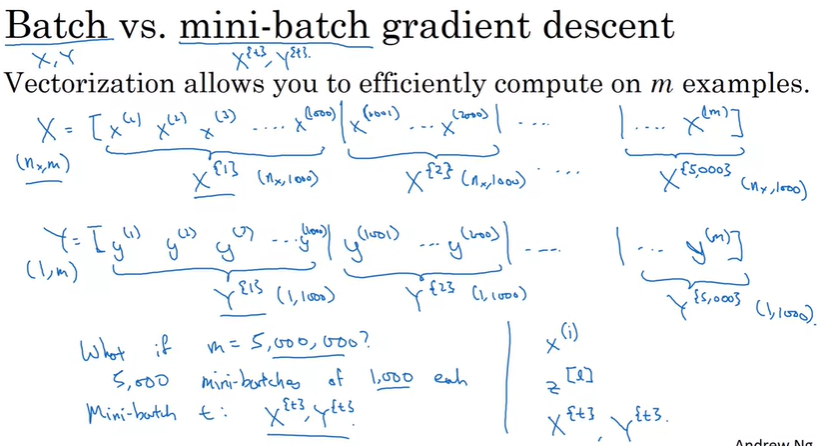
Mini-batch gradient descent

thay về process toàn bộ data của tập train và test (facebook vs cả tỉ)

thì mình chỉ process một phần của tập (mini-batch) để cải thiện tốc độ train



note nhỏ:

x(i) là train thứ i

z[l] là layer thứ l

X{t}, Y{t} là mini - batch thứ t

cách triển khai

Text, letter

Description automatically generated

note 1 epoch là một lần pass của một tập (hay là cả cập?) mini – batch

vấn đề tiếp là chọn mini – batch size sao cho phù hợp

Diagram, text

Description automatically generated

việc chọn to quá (VD: = chính size của tập train) thì chả khác gì train toàn tập (hiện đang chậm do quá nhiều)

việc chọn nhỏ quá (VD: = 1) thì train có khi lâu hơn, vì k có vectorize, kèm theo đó là siêu nhiễu (mỗi thằng một phách)

trong mức vừa thì vừa cải thiện “hơn” tốc độ train, và ít noise hơn

Exponentially weighted averages

Chart, scatter chart

Description automatically generated

hiểu nôm na như kiểu trải phẳng hóa, trung bình lại cái đồ thị

biến số mới là V giờ sẽ “phẳng” hay “sóc” phụ thuộc vào Beta

như ông Andrew nói V sẽ là trung bình của 1 / (1 – beta) ngày

do vậy nên nếu

B lớn thì số ngày nhiều đồ thị là đường xanh lá sẽ mịn hơn

nhưng khó adapt vs biến đổi bất thường của thời tiết

B vửa như 0.9 sẽ là đường màu đỏ trông khá hay và vừa vừa

còn nếu B nhỏ quá sẽ càng “sóc” vì giờ chỉ là trung bình của 1-2 ngày nên sẽ “sóc”

nhưng sẽ adapt tốt với biến đối của thời tiết

Diagram

Description automatically generated with medium confidence

qua hình ảnh trên cái exponecially này là tính trung bình

theo kiểu càng gần ngày hiện tại thì trọng số càng lớn

càng xa ngày hiện tại thì trọng số càng nhỏ

cái này tức là kiểu cho thấy xu hướng sát với hiện tại nhất

bias correction

Chart, scatter chart

Description automatically generated

ở hình ảnh trên một vấn đề nhỏ của exponential weight average là tại những điểm ban đầu

do chưa có data trước ngày bắt đầu VD: ngày 1, 2, 3, …

nếu ốp công thức vào thì giá trị sẽ rất thấp VD: V1 = 0 + 0.02\*theta1 (tức sai)

(do bản chất là trung bình của 50 ngày mà chỉ có 1 nên sẽ khá là sai)

do vậy cần correct lại những giá trị init = cách chia cho 1 – Betat để

“scale” lại cho đúng là trung bình của các ngày có thể

Gradient descent with momentum

Diagram

Description automatically generated

idea có thể thấy được là việc training có thể có nhiễu những một đặc điểm của grad decent

là luôn có xu hướng về cực tiểu, tận dụng điều này liệu ta có thể bằng một cách nào đó

thêm một đại lượng giống như một lực, ép lên biến hiện tại của ta?

đúng vậy hoàn toàn có thể giống như exponetial average làm “mềm”

và các giá trị trước đó đều đóng vai trò trog giá trị hiện tại

thì trong trường hợp này ta có thể áp dụng vào grad để khiến việc tích lũ gia tốc, lực

qua các lẫn train cũ sẽ có thể cải thiện tốc độ train

“hình dung như một viên bi, với tình huống k có tác động nó lăn lung tung nhưng vẫn có hướng về đáy quả cầu”

“nhưng nếu như tác dụng vào nó một lực có chủ đích, độ lung tung sẽ giảm và nó sẽ lăn nhanh hơn”

Text, letter

Description automatically generated

RMSprop

cũng là một dạng tăng tốc về đích nhưng cách tính khác nhau bằng cách nhân, chia vs learning rate

thì cũng khiến tăng giảm biến động của trọng số qua đó tăng tốc về đích

lưu ý nhỏ duy nhất là có thể dính chia cho 0 thì có thể cộng vs epsilon nhỏ khoảng 10-8 là đủ

Adam optimization algorithm

Text, letter

Description automatically generated

adam đã combine cả momentum và rmsprop vào và biến thành một racing boy trong làng train model

Learning rate decay

Diagram

Description automatically generated

một trọng số mà cũng khá quan trọng đó chính là learning rate vì learning rate

quá cao quá nhỏ thì đã biết rồi

nhưng nó cũng cần phải thay đổi trong quá trình train

vì có thể khi gần tới đích nó lại cao quá làm vượt hỏi cực tiểu

hoặc nhỏ quá thì lại k vượt qua đc cực tiểu địa phương

Chart, diagram

Description automatically generated

vấn đề cực tiểu địa phương và yên ngựa, và điểm cực thoải kiểu “đồng bằng” thì train siêu lâu nữa

Diagram

Description automatically generated