What is Loss Function?

23 March 2022 07:15

Loss function is a method of evaluating how well your algorithm is modelling your dataset.

high → poor

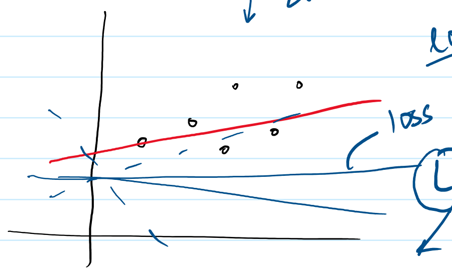
small → great

$f(x) = x^2 + 2$

L(parameters)

↓ 2var

loss/error minimise



(m,b)

$L = (y_i - \hat{y}_i)^2$

$\hat{y}_i = mx_i + b$

loss function eye of ML algorithm

GD

$L(m,b) = \min$

Why is Loss function important?

[You can't improve what you can't measure.]

Peter Drucker

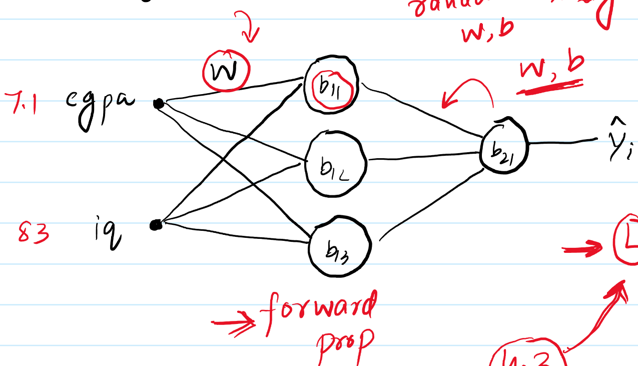
Loss Function in Deep Learning?

cgpa	iq	package(lpa)
7.1	83	3.2
8.5	91	4.5
6.3	102	6.1
5.1	87	2.7
..	..	..
..	..	..
..	..	..

Backprop

multiple epochs

random w, b → gradient descent



(3.7)

$L = (y_i - \hat{y}_i)^2$

$(3.2 - 3.7)^2$

4.3

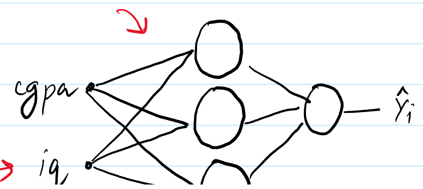
→ forward prop

Loss functions in DL

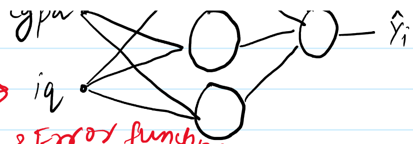
Keras

- Object detection
  - focal loss
- Regression
  - mse
  - mae
  - huber loss
- Classification
  - binary cross entropy
  - categorical cross entro
  - hinge loss
- Autoencoders
  - KL divergence
- GAN
  - discriminator loss
  - minmax gan loss
- Embedding
  - Triplet loss

Loss Function Vs Cost Function



Loss Function Vs Cost Function



Error function

Loss function → single training eg

$y_i = 6.3 \quad \hat{y}_i = 6.1$

$(y_i - \hat{y}_i)^2$

$(6.3 - 6.1)^2 =$

batch Cost function

$\frac{1}{n} \sum (y_i - \hat{y}_i)^2$

$\frac{1}{4}\left[(6.1-6.3)^2 + (4.1-4)^2 + (3.5-3.7)^2 + (7.2-7)^2\right] = CF$

5 lpa $\sqrt{5}$ lpa $(lpa)^2$

cgpa	iq	package (y_i)	Prediction ($\hat{y}_i$)	error $(y_i - \hat{y}_i)^2$
6.3	100	6.3	6.1	0.2
7.1	91	4.1	4	0.1
8.5	83	3.5	3.7	−0.2
9.2	102	7.2	7	0.2
8.1	100	50	8	

overall error

1. Mean Squared Error (MSE)

squared loss L2 loss

$(y_i - \hat{y}_i)^2$

$(\text{true} - \text{predict})^2$

$(6.3 - 6.1)^2 =$

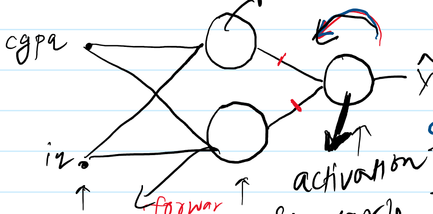
$\boxed{(y_i - \hat{y}_i)^2}$

Advan DISAdg

mae

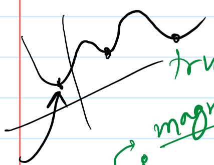
$|y_i - \hat{y}_i|^2$

quadratic $\wedge 2$ punish



$(42)^2$

drastic big



true − predic → 1 unit → 1 unit

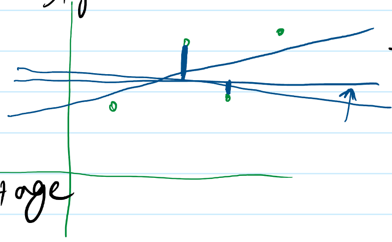
2 unit → 4 unit

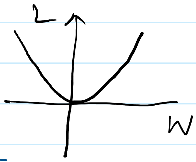
4 unit → 16 unit

magnify

$(y_i - \hat{y}_i)^2$

$(w, b) \to L$ min





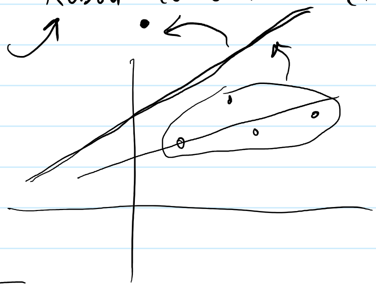
Advantages

1) Easy to intreprt
2) Diffrentiable (GD)
3) 1 local minima

$$CF = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Disadvantage

1) Error unit (squared) → diff
2) Robust to Outliers (Not)



mse

2. Mean Absolute Error (MAE) → L1 Loss

$L = |y_i - \hat{y}_i|$

2 | abs

$C = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|^2$

$|\text{true} - \text{predict}|$

punish

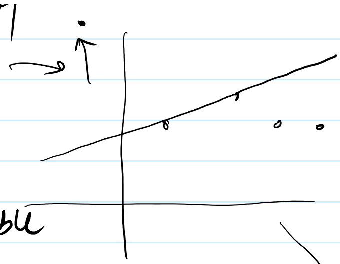
2 lpa

cgpa | iq | pack lpa $|(y_i - \hat{y}_i)|$

lpa

$C = \frac{1}{n}\sum_{i=1} |y_i - \hat{y}_i|$ | true − predict |

punish



Advantages

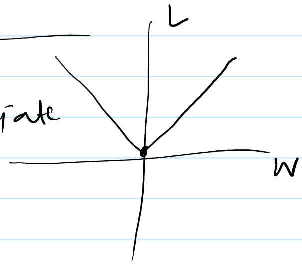
1) Intuitive and easy
2) Unit → same − y
3) Robust to outliers

Disadvange

1) Not differentiable

GD → differentiate

Subgradient



mae huber mse

3. Huber Loss ✓

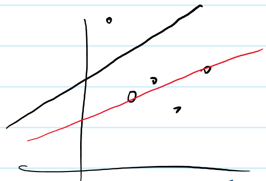
$$L = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{for } |y-\hat{y}| \le \delta \\ \delta|y-\hat{y}| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}$$

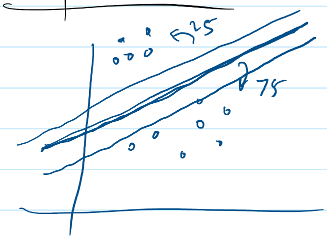
mse

hyper

mae

mse − outliers ✓
mae − normal point ✓





log loss

4. Binary Cross Entropy

→ classification ✓
→ Two classes ✓

100 days

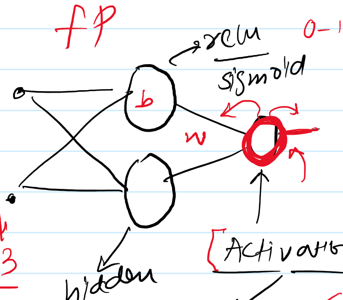
cgpa	iq	placement
8	80	1
7	70	0
6	60	0

0.12

→ 1 0

GD FP

0.73
0.25
0.13



[Activation]

{sigmoid}

Loss function $= \boxed{-y\log(\hat{y}) - (1-y)\log(1-\hat{y})}$

Keras

$y \rightarrow$ actual value / target

$\hat{y}$ − NN prediction

0.12 $-(1-0)\log(1-0.25)$

$-1\log(0.75)$ −0.12

$-1\log(0.73)$

$-1\times -0.13 = 0.13$

cost function $= -\frac{1}{n}\left[\sum_{i=1}^{n} y_i\log\hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right]$

maximum likeli

[Logistic Reg]

Advantage

* Differentiable

Disadvan

→ multi local minima
→ Intuitive

5. Categorical Cross Entropy [used in Softmax Regression]

OHE

cgpa | iq | placed? | Yes | No | Maybe

→ [Multi-class classification] ✓

5. Categorical Cross Entropy [used in softmax Regression]

→ [Multi-class] {classification}

1 point

$$L = -\sum_{j=1}^{k} y_j \log(\hat{y}_j)$$

cgpa	iq	placed?
8	80	Yes 1
6	60	No 2
7	70	Maybe 3

Yes	No	Maybe	
1	0	0	Yes
0	1	0	No
0	0	1	Maybe

where (k) is # classes in the data ↘ 3

1 point

$$L = -y_1 \log(\hat{y}_1) - y_2 \log(\hat{y}_2) - y_3 \log(\hat{y}_3)$$

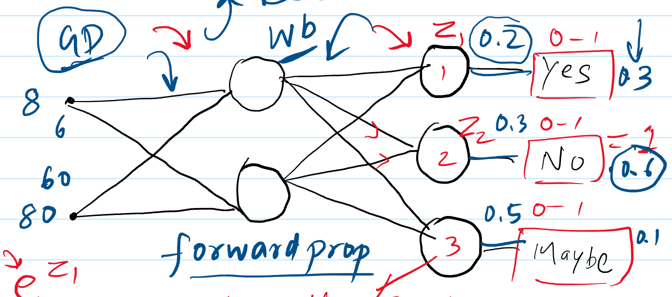
[0.2 0.3 0.5]

[1 0 0]

$L = -1 \times \log(0.2)$

$\frac{e^{z_3}}{e^{z_1}+e^{z_2}+e^{z_3}}$ $\frac{e^{z_2}}{e^{z_1}+e^{z_2}+e^{z_3}}$

$$f(z) = \frac{e^{z_1}}{e^{z_1}+e^{z_2}+e^{z_3}}$$ softmax



class ← categories

[0.3 0.6 0.1]

[0 1 0]

$-1 \times \log(0.6) = -$

Sparse Categorical Cross Entropy

cgpa	iq	placed	OHE
7	70	Yes	1
8	80	No	2
6	60	Maybe	3

1/2/3

[0.1 0.4 0.5]

L ↘ $-1 \times \log(0.1)$

[0.1 0.4 0.6]

$-1 \times \log(0.4)$

[0.6 0.2 0.2]

$-1 \times \log(0.2)$

{SCE, CCE} Loss function

SCE ↘ fast

fast

OHE — fast

SCE

$$\boxed{L = -\sum_{j=1}^{k} y_j \log(\hat{y}_j)}$$ ↘ Cost function

$$\boxed{C = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} y_{ij} \log(\hat{y}_{ij})}$$

- Reg
  - mse
  - Outliers — mae
    - huber loss
- Classification
  - binary → bce
  - multi
    - CCE
    - SCE