15/7/2021

Binary Cross Entropy / log loss

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(p(y_i)) + (1-y_i)\cdot \log(1-p(y_i))$$

cross entropy

$$H_p(q) = -\sum_{c=1}^{C} q(y_c)\log(p(y_c))$$

~~Los~~

cross entropy − Entropy >= 0

Kullback-Liebler Divergence

→ measure of dissimilarity between two distributions

$$D_{KL}(q\|p) = H_p(q) - H(q)$$

$$= \sum_{c=1}^{C} q(y_c)\cdot[\log(q(y_c)) - \log(p(y_c))]$$

Understanding optimizers

→ goal to maximize/minimize the value of an objective function w.r.t parameters of that function.

Scanned with CamScanner

Page No.:
Date:

1.) Gradient Descent

$$x_{i+1} = x_i - \alpha . \nabla F(x_i)$$

In NN,
$\theta_0 \rightarrow$ single vector for initial weights and bias
$D \rightarrow$ entire dataset

$$F(\theta) = \frac{1}{|D|} \sum_{x \in D} f(x|\theta)$$

$f(x|\theta) \rightarrow$ loss for a single example x when using model parameter θ.

2.) Stochastic gradient descent

Since computing the $\nabla F(\theta)$ for every step is impractical, we approximate this gradient using smaller mini batch of data

$$\nabla F(\theta) \approx \nabla \tilde{F}_i(\theta) = \frac{1}{|B_i|} \sum_{x \in B_i} \nabla f(x|\theta)$$

where B_i batches partition into the dataset into smaller subsets (of equal size)

$$\therefore \quad x_{i+1} = x_i - \alpha \nabla F(x_i)$$

$$\Downarrow$$

$$\theta_{i+1} = \theta_i - \alpha . \nabla \hat{F}_i(\theta_i)$$

Scanned with CamScanner

Page No.:
Date:

$$\text{Accuracy} = \frac{\text{Number of correct prediction}}{\text{Total number of prediction}}$$

For binary classification

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

TP = True positive
TN = " Negative
FN = False ~~negative~~ positive
FN = " Negative

Some preprocessing ⇒

(i) min max scaling

$$m = \frac{X - X_{min}}{X_{max} - X_{min}}$$

m → new value
x → original cell value
x_{min} → min value of column
x_{max} → max value of column

Page No.:
Date:

Some examples:

i) Binary cross entropy

idx	feature			w_sum	target y	prediction $\hat{y}$	loss*
0	0.1	0.5	0.3	0.26	1	0	0.58
1	0.2	0.3	0.1	0.1	0	0	0.09
2	0.7	0.4	0.2	0.2	1	0	0.32
3	0.1	0.4	0.3	0.3	0	0	0.15

* $$\text{loss} = -(y \times \log(\text{w\_sum})) + (1-y)\log(1-\text{w\_sum}))$$

w_sum = weighted sum

$$\text{average} = \frac{\Sigma \text{ loss}}{n}$$

$$= \frac{1.14}{4} = \underline{\underline{0.285}}$$

Scanned with CamScanner