Cross-entropy (loss)

To find cross-entropy loss, we need 2 distributions

These can be any 2 distⁿ. Generally in ML,

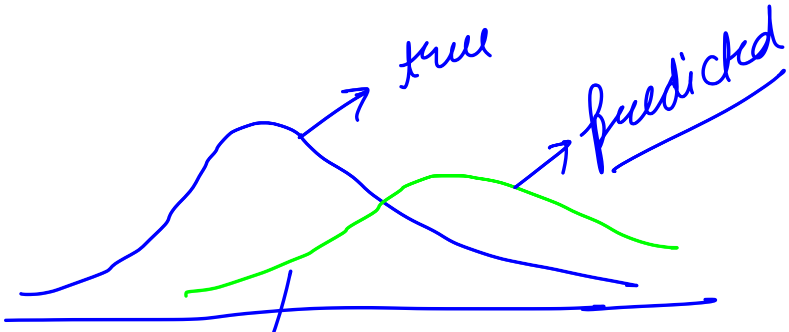
these 2 distⁿ are :-

i) true distⁿ of a training sample (obtained from true y)

ii) predicted distⁿ of the same training sample (obtained from the model)

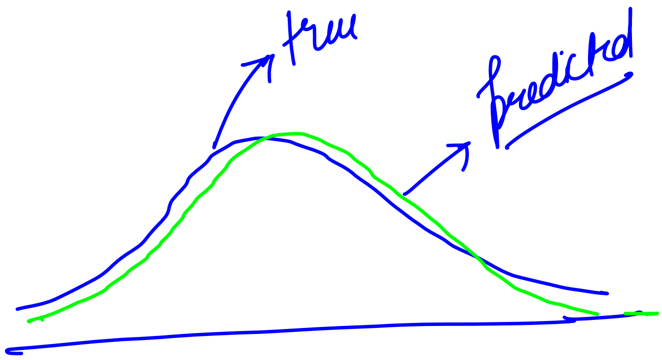
So, we find the cross-entropy b/w (i) & (ii)

→ this value of cross-entropy gives an indication of how close the predicted distⁿ is to the true distⁿ



not close to true distⁿ

⇒ high cross-entropy loss



low cross-entropy loss

Math

Cross entropy for one training example $= -\log_2$ [prob. of true class as predicted by the model]

I find this descriptive way easy to remember & explain

Lets find cross-entropy loss for one example

(Problem 0, hw4)

Model: logistic regⁿ, $w = (1, -30, 3)$

Dataset:

	f_1	f_2	f_3	y
sample 1 → x_1	20	0	0	1
sample 2 → x_2	23	1	1	0

features

for sample 1, true class = 1

for sample 2, true class = 0

Cross-entropy loss for sample 1

For sample 1, true class = 1 , so $\text{loss} = -\log\left[p(Y=1)\right]$

What else do we need?

→ prob. of true class as predicted by model
$\qquad\qquad\quad\downarrow$
which is 1

$$p(Y=1) = ??$$

logistic regⁿ

$$p(Y=1) = \frac{1}{1+e^{-\cancel{x}_1 w}}$$

$$= \frac{1}{1+e^{-(20\ 0\ 0)\cdot\begin{bmatrix}1\\-30\\3\end{bmatrix}}}$$

$$= \frac{1}{1+e^{-20}} = 0.999 \simeq 1$$

So, true class = 1

prob. of $(Y=1)$ = prob. of true class as predicted by the model $\simeq 1$

⇒ our model is really good at predicting for this sample

$\Rightarrow$ loss should be less

$$\text{loss} = -\log_2(1) \simeq \underline{\underline{0}} \checkmark$$

<u>Cross - entropy loss for sample 2</u>

true class for sample 2 is $\underline{\underline{0}}$. $(y=0 \text{ for } x_2)$

So, $\text{loss} = -\log_2\left[p(Y=0)\right]$

$$p(y=0)_{\text{logistic reg}^n} = \text{??}$$

* <u>Imp.</u> $\rightarrow$ logistic reg^n gives the prob. $\underline{\underline{p(Y=1)}}$

$$p(Y=1)_{\text{for } x_2} = \frac{1}{1+e^{-x_2 w}}$$

$$= \frac{1}{1+e^{-(23 \quad 1 \quad 1)\begin{bmatrix}1\\-30\\3\end{bmatrix}}}$$

$$= \frac{1}{1+e^{4}}$$

$$= \frac{1}{55.6} = 0.018$$

$$\therefore p(Y=0) = 1 - 0.018 = 0.982$$

$$\text{loss} = -\log_2 (p(Y=0))$$

$$= -\log_2 (0.982) = 0.026$$

$\therefore$ Total loss from 2 training samples

$$= \underbrace{0}_{\text{loss for sample 1}} + \underbrace{0.026}_{\text{loss for sample 2}} = 0.026$$

<u>Mistake that most groups made in this ques.</u>

→ Both groups found the predicted diffⁿ for 2 samples

and found the loss b/w those 2 diffⁿ

predicted distⁿ

	$p(y=1)$	$p(y=0)$
sample 1	1 (or 0.999)	0
sample 2	0.018	0.982

$$\text{loss} = -[1.\log(0.018) + 0.\log(0.982)]$$

$$= 5.796 \rightarrow \text{which is wrong}$$

wrong because both of them are predicted distⁿ

what should've been done

predicted distⁿ

	$p(y=1)$	$p(y=0)$
sample 1	1 (or 0.999)	0
sample 2	0.018	0.982

true distⁿ

$y=1$	$y=0$
1	0
0	1

loss = − p log q

true distⁿ

predicted distⁿ

this is just a more mathematical way of saying loss = − log [prob. of true distⁿ as predicted by the model]

using this

Sample 1

predicted

$$\text{loss} = -\Big[\big(\underbrace{1 \log(0.999)}_{\approx 0} + \cancel{0 \log 0}_{=0}\big) + \big(\cancel{0 \log(0.018)}_{=0} + 1 \cdot \log(0.982)\big) \Big]$$

true

Sample 2

true

predicted

$$= -\log_2(0.982) = 0.026$$