

Today: implementing neural LMs
in PyTorch

→ cross entropy loss

- used in NLMS, as well
as other classification tasks

What is a loss fn?

↳ intuitively, tells us how bad a model
is doing at predicting the training data

↳ in NLMS, how bad is the model
at predicting the next word

assume we have a training ex

"students opened their"

⇒ books

└──────────────────┘

model input

└──────────┘

desired model
prediction

$P(\text{"books"} \mid \text{"students opened their"})$

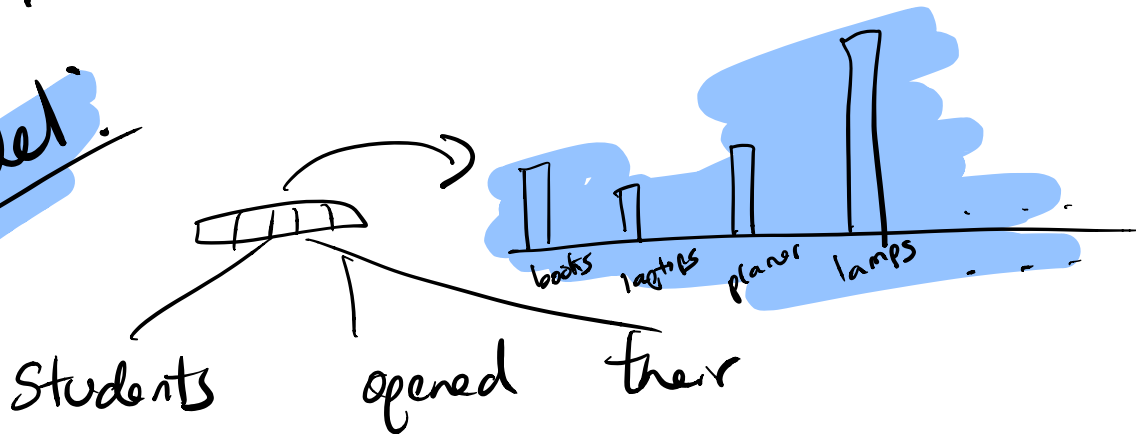
↳ maximize this probability!

We want to minimize the negative log probability of "books"

$$L = -\log(p(\text{books} | \text{students opened their}))$$

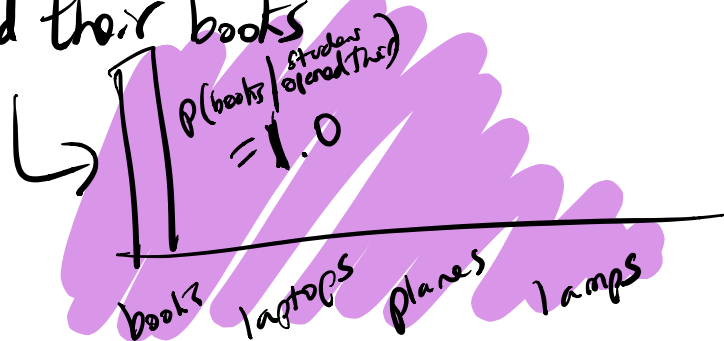
Why is this called the cross-entropy loss?

Model:



data:

Students opened their books



Cross entropy of two distributions p and q

\Rightarrow quantifies distance between the distributions

$$H(p, q) = - \sum_{w \in V} p(w) \log q(w)$$

↑
1 for "books"
0 every other word
type

↑
predicted
conditional
probs

$$= - \log q(\text{books} | \dots)$$