50.021 -AI

Alex

Week 03: Basics of neural networks

suppose we have a layer which computes $Y = f(X)$ for some function f.

The code uses DY in the backwardpass, this is $\frac{\partial E}{\partial Y}$ - where Y the outputs of the current layer. This is the gradient that comes from above

We need to do two things:

- if f has learnable parameters W, then: compute $\frac{dE}{dW} = DY * \frac{\partial f}{\partial W}$, store this as $self.DW$ – this needs NOT to be done for the tanh layer, but for $self.DB$, $self.DW$ in the linear layer
- always we have to compute

$$\frac{dE}{dX} = DY * \frac{\partial f}{\partial X}$$

and let the backward function return this term $\frac{dE}{dX}$