

TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2020

Backpropagation with Arrays and Tensors

Handling Arrays

$$\begin{aligned} \mathbf{h} &= \sigma \left(W^0 \mathbf{x} - B^0 \right) \\ \mathbf{s} &= \sigma \left(W^1 \mathbf{h} - B^1 \right) \\ P_{\Phi}[\hat{y}] &= \underset{\hat{y}}{\text{softmax}} \mathbf{s}[\hat{y}] \\ \mathcal{L} &= -\ln P[y] \end{aligned}$$

Each array (matrix) \mathbf{W} is represented by an object with attributes $\mathbf{W.value}$ and $\mathbf{W.grad}$.

$\mathbf{W.grad}$ is an array storing $\nabla_{\mathbf{W}} \mathcal{L}$.

$\mathbf{W.grad}$ has same indices (same “shape”) as $\mathbf{W.value}$.

Source Code Loops

$$s = \sigma(W h - B)$$

Can be written as

$$\text{for } j \quad \tilde{s}[j] = 0$$

$$\text{for } j, i \quad \tilde{s}[j] += W[j, i] h[i]$$

$$\text{for } j \quad s[j] = \sigma(\tilde{s}[j] - B[j])$$

Backpropagation on Loops

the backpropagation for

$$\text{for } j \text{ } \textcolor{red}{s[j]} = \sigma(\tilde{s}[j] - B[j])$$

is

$$\text{for } j \text{ } \tilde{s}.\text{grad}[j] \textcolor{red}{+=} \textcolor{red}{s}.\text{grad}[j] \sigma'(\tilde{s}[j] - B[j])$$

$$\text{for } j \text{ } \textcolor{red}{B}.\text{grad}[j] \textcolor{red}{-=} \textcolor{red}{s}.\text{grad}[j] \sigma'(\tilde{s}[j] - B[j])$$

Backpropagation on Loops

the backpropagation for

$$\text{for } j, i \quad \tilde{s}[j] \ += \ W[j, i]h[i]$$

is

$$\text{for } j, i \quad W.\text{grad}[j, i] \ += \ \tilde{s}.\text{grad}[j]h[i]$$

$$h.\text{grad}[i] \ += \ \tilde{s}.\text{grad}[j]W[j, i]$$

General Tensor Operations

In practice all deep learning source code can be written using scalar assignments and loops over scalar assignments. For example:

$$\begin{aligned} \text{for } h, i, j, k \quad \tilde{Y}[h, i, j] &+= A[h, i, k] B[h, j, k] \\ \text{for } h, i, j \quad Y[h, i, j] &= \sigma(\tilde{Y}[h, i, j]) \end{aligned}$$

has backpropagation loops

$$\begin{aligned} \text{for } h, i, j \quad \tilde{Y}.\text{grad}[h, i, j] &+= Y.\text{grad}[h, i, j] \sigma'(\tilde{Y}.\text{grad}[h, i, j]) \\ \text{for } h, i, j, k \quad A.\text{grad}[h, i, k] &+= \tilde{Y}.\text{grad}[h, i, j] B[h, j, k] \\ \text{for } h, i, j, k \quad B.\text{grad}[h, j, k] &+= \tilde{Y}.\text{grad}[h, i, j] A[h, i, k] \end{aligned}$$

END