TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2020

Backpropagation with Arrays and Tensors

Handling Arrays

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

Each array (matrix) W is represented by an object with attributes W.value and W.grad.

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

W.grad has same indeces (same "shape") as W.value.

Source Code Loops

$$s = \sigma\left(W^1 h - B^1\right)$$

Can be written as

$$\begin{array}{lrcl} \text{for } j & \tilde{s}[j] & = & 0 \\ \text{for } j, i & \tilde{s}[j] & \mathrel{+}= & W^1[j,i]h[i] \\ \text{for } j & s[j] & = & \sigma(\tilde{s}[j] - B^1[j]) \end{array}$$

Backpropagation on Loops

the backpropagation for

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

is

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

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

Backpropagation on Loops

the backpropagation for

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

is

$$\text{for } j,i \;\; W.\mathrm{grad}[j,i] \mathrel{\texttt{+=}} \tilde{s}.\mathrm{grad}[j]h[i]$$

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

General Tensor Operations

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

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

has backpropagation loops

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

END