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

$$s = \sigma\left(Wh - B\right)$$

Can be written as

$$\begin{aligned}
&\text{for } j \quad \tilde{s}[j] \;=\; 0 \\
&\text{for } j, i \;\; \tilde{s}[j] \;\texttt{+=}\; W[j,i]h[i] \\
&\text{for } j \quad s[j] \;=\; \sigma(\tilde{s}[j] - B[j])
\end{aligned}$$

Backpropagation on Loops

the backpropagation for

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

is

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

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

Backpropagation on Loops

the backpropagation for

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

is

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

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

General Tensor Operations

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

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

has backpropagation loops

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

END