Flipped Classroom Questions on Computational Graphs, and Backpropagation

Michael Collins

Question 1: Consider the following system of equations, which define a neural network with two hidden layers:

Definitions: The set of possible labels is $\mathcal{Y}$. We define $K = |\mathcal{Y}|$. $g^1 : \mathbb{R}^m \rightarrow \mathbb{R}^m$ and $g^2 : \mathbb{R}^m \rightarrow \mathbb{R}^m$ are transfer functions. We define LS = LOG-SOFTMAX.

Inputs: $x^i \in \mathbb{R}^d, y^i \in \mathcal{Y}$, $W^1 \in \mathbb{R}^{m \times d}$, $b^1 \in \mathbb{R}^m$, $W^2 \in \mathbb{R}^{m \times m}$, $b^2 \in \mathbb{R}^m$, $V \in \mathbb{R}^{K \times m}$, $\gamma \in \mathbb{R}^K$.

Equations:

$$\begin{aligned}
z^1 \in \mathbb{R}^m &= W^1 x^i + b^1 \\
h^1 \in \mathbb{R}^m &= g^1(z^1) \\
z^2 \in \mathbb{R}^m &= W^2 h^1 + b^2 \\
h^2 \in \mathbb{R}^m &= g^2(z^2) \\
l \in \mathbb{R}^K &= V h^2 + \gamma \\
q \in \mathbb{R}^K &= \text{LS}(l) \\
o \in \mathbb{R} &= -q_{y_i}
\end{aligned}$$

Question 1a: Draw the computational graph for this system of equations. Which variables are at the leaves?

Question 1b: Write down expressions (using products of Jacobians) for the following quantities:

$$\left.\frac{\partial o}{\partial V}\right|^{\bar{f}^o}$$

$$\left.\frac{\partial o}{\partial W^1}\right|^{\bar{f}^o}$$

$$\left.\frac{\partial o}{\partial W^2}\right|^{\bar{f}^o}$$

Question 2: Consider a computational graph with the following definitions:

- Number of leaves $l = 2$, number of nodes $n = 5$
- A variable $u^i \in \mathbb{R}$ for $i = 1 \dots n$. Hence each variable in the graph has dimension $d_i = 1$
- Set of edges $E = \{(1,3), (2,3), (1,4), (2,4), (3,5), (4,5)\}$
- Local functions:
$$u^3 = f^3(u^1, u^2) = u^1 \times u^2$$
$$u^4 = f^4(u^1, u^2) = u^1 + u^2$$
$$u^5 = f^5(u^3, u^4) = 2 \times u^3 \times u^4$$

Question 2a: Draw the computational graph for this example. Assume that the inputs are
$$u^1 = 3, \quad u^2 = 4$$
Show how values are computed in the forward pass of the algorithm, giving an output value for u^5

Question 2b: The output value u^n is a function $\bar{f}^n$ of the values for the leaf variables u^1 and u^2.
$$u^n = \bar{f}^n(u^1, u^2)$$
Write down the expressin for $\bar{f}^n$

Question 2b: Recall that for each edge (j, i) we define the Jacobian
$$J^{j \to i}(A^i) = \left. \frac{\partial u^i}{\partial u^j} \right|^{f^i}$$
where $A^i = \langle u^j : (j,i) \in E \rangle$. For example we have
$$J^{1 \to 3}(A^3) = \frac{\partial}{\partial u^1} \left(u^1 \times u^2 \right) = u^2$$
Write down expressions for Jacobians associated with the other edges in graph. Calculate the values for these Jacobians under inputs $u^1 = 3$ and $u^2 = 4$

Question 2c: Recall that the general form for the backward pass is:

- $p^n = 1$

- For $j = (n-1)\dots 1$:

$$p^j = \sum_{i:(j,i)\in E} p^i J^{j\to i}(A^i)$$

Given that the inputs are $u^1 = 3$ and $u^2 = 4$, calculate the values $p^5, p^4, \dots p^1$ calculated in the backward pass.

Question 3: Consider the following system of equations, which define a neural network with a single hidden layer:

Definitions: The set of possible labels is $\mathcal{Y}$. We define $K = |\mathcal{Y}|$. $g : \mathbb{R}^m \to \mathbb{R}^m$ is a transfer function. We define LS = LOG-SOFTMAX.

Inputs: $x^i \in \mathbb{R}^d, y^i \in \mathcal{Y}, W \in \mathbb{R}^{m\times d}, b \in \mathbb{R}^m, V \in \mathbb{R}^{K\times m}, \gamma \in \mathbb{R}^K$.

Equations:

$$\begin{aligned} z \in \mathbb{R}^m &= Wx^i + b \\ h \in \mathbb{R}^m &= g(z) \\ l \in \mathbb{R}^K &= Vh + \gamma \\ q \in \mathbb{R}^K &= \text{LS}(l) \\ o \in \mathbb{R} &= -q_{y_i} \end{aligned}$$

Question 3a: Now say we have a pair of training examples $(x^{i,1}, y^{i,1})$ and $(x^{i,2}, y^{i,2})$, and we would like to take gradients with respect to the loss function

$$L(\theta, v) = -\log p(y^{i,1}|x^{i,1}; \theta, v) - \log p(y^{i,2}|x^{i,2}; \theta, v)$$

Write down a system of equations for this loss function. Show the computational graph. Your graph should have intermediate variables $z^1, z^2, h^1, h^2, l^1, l^2, q^1, q^2$, and an output variable o.

Question 3b: If o is the output variable for your answer to Question 3a, write down an expression for

$$\frac{\partial o}{\partial V}\bigg|^{\bar{f}^o}$$

Hint: recall that to calculate a partial derivative of the output with respect to a leaf, you can sum over directed paths from the leaf to the output, and take the product of Jacobians along each path.

Question 3c: Again assume that we have inputs $(x^{i,1}, y^{i,1})$ and $(x^{i,2}, y^{i,2})$ and we would like the loss function to be

$$L(\theta, v) = -\log p(y^{i,1}|x^{i,1}; \theta, v) - \log p(y^{i,2}|x^{i,2}; \theta, v)$$

Assume that we would like to implement this loss through the following system of equations:

$$\begin{aligned} z \in \mathbb{R}^{m\times 2} &= f^z(W, x^{i,1}, x^{i,2}, b) \\ h \in \mathbb{R}^{m\times 2} &= f^g(z) \\ l \in \mathbb{R}^{K\times 2} &= f^l(V, h, \gamma) \\ q \in \mathbb{R}^{K\times 2} &= f^q(l) \\ o \in \mathbb{R} &= f^o(q, y^{i,1}, y^{i,2}) \end{aligned}$$

How would you define the functions f^z, f^g, f^l, f^q, and f^o to implement the loss function?

You may find the following notation useful. Given matrices $A \in \mathbb{R}^{m\times d_1}$ and $B \in \mathbb{R}^{m\times d_2}$, we write $[A; B]$ to refer to the matrix of dimension $m \times (d_1 + d_2)$ formed by concatenating the columns of B to the columns of A.