1. Consider function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = \log(\sum_{i=1}^{n} \exp(x_i))$$

This function is called log-sum-exponential. Find the gradient $\nabla f(x)$. (Does the answer look familiar?)

2. Suppose softmax : $\mathbb{R}^n \to \mathbb{R}^n$ be the softmax function. Also let $\mathbf{e}_i$ be the unit vector in $\mathbb{R}^n$ whose i-th element is 1 and others are zero (one-hot vector). Consider the cross-entropy loss $L : \mathbb{R}^n \to \mathbb{R}$ such that

$$L(x) = -\mathbf{e}_i^T \cdot \log[\text{softmax}(x)]$$

where log is applied in the elementwise manner. Find gradient $\nabla L(x)$.

3. Consider function $f(W) : \mathbb{R}^{n \times n} \to \mathbb{R}$ such that

$$f(W) = x^T W x$$

Here the variable W is $n \times n$ matrix. x is n-dimensional constant vector. (This is called the quadratic form of x using matrix W.) Find Jacobian

$$\frac{df}{dW}$$

where log is applied in the elementwise manner. Find gradient $\nabla L(x)$.

4. We have computational graph in Fig 1. The numbers above the flows represent the forward values. The numbers below the flows represent the gradient with respect to f. Fill in (a)–(d).

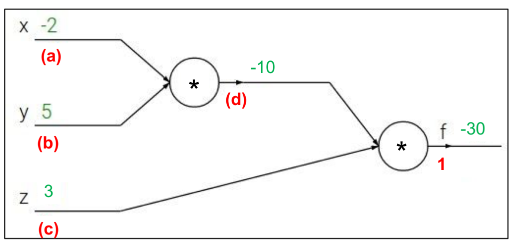

Figure 1: computational graph

5. Consider loss function $L(x, y) = 2x + 3xy$ with learning rate α. We would like to minimize the loss using gradient descent. What is the step for gradient descent at point (x, y)?

6. Consider a convolutional layer. The input map has shape $(3, 32, 32)$. That is, there are 3 channels, and the map width and height is 32. Now the filter has width and height of 3, and the depth is 3. The number of filters is given by 8. The stride is 1. What is the shape of the output? Write the shape as

$$(\text{number of filters}, \text{output map width}, \text{output map height})$$