Problem Set #1

Simas Glinskis

January 11, 2017

In collaboration with Natasha Antropova

Problem 1

a

Using chain rule,

$$\frac{\partial l}{\partial x} = \frac{\partial l}{\partial u}\frac{\partial u}{\partial x} = \frac{\partial l}{\partial u}\frac{\partial f}{\partial x} = \frac{\partial l}{\partial u}(\frac{\partial l}{\partial y}\frac{\partial g}{\partial x} + \frac{\partial l}{\partial z}\frac{\partial h}{\partial x})$$

This matches the computation in the slides, x.grad = u.grad(y.grad$\frac{\partial g}{\partial x}$+z.grad$\frac{\partial h}{\partial x}$).

b

Here, we set u.grad = 1, to make the computation cleaner.

$$\frac{\partial}{\partial x}(\frac{\partial l}{\partial x}) = \frac{\partial}{\partial x}(\frac{\partial l}{\partial y}\frac{\partial g}{\partial x} + \frac{\partial l}{\partial z}\frac{\partial h}{\partial x})$$

$$\frac{\partial^2 l}{\partial x^2} = \frac{\partial^2 l}{\partial x \partial y}\frac{\partial g}{\partial x} + \frac{\partial l}{\partial y}\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 l}{\partial x \partial z}\frac{\partial h}{\partial x} + \frac{\partial l}{\partial z}\frac{\partial^2 h}{\partial x^2}$$

Already we see that this cannot simplify to the expression in the homework, due to the cross terms which leave first order derivatives.

Problem 2

Softmax is defined as,

$$P(x)_i = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

Let us call the EdF Softmax class P, the derivative for the backward method is,

$$\frac{\partial l}{\partial x} = \frac{\partial l}{\partial P}\frac{\partial P}{\partial x}$$

where,
$$\frac{\partial P}{\partial x} = \frac{e^x}{\sum_j e^{x_j}} - (\frac{e^x}{\sum_j e^{x_j}})^2 = P(x)(1 - P(x))$$
Therefore,
$$\frac{\partial l}{\partial x} = \frac{\partial l}{\partial P} P(x)(1 - P(x))$$
Or, in code, x.grad = self.grad*self.value(1.0-self.value), which is what we have in the edf.py code.

Problem 3

See notebook.