Homework 03: Jacobians and the application of Chain Rule.

Yann LeCun
The Courant Institute, New York University

This problem set is designed to practice the application of chain rule and the differentiations of various multivariate functions. This is what you need to do to write the `bprop` method of a module.

If you know how to compute the derivatives of simple functions, you have all the skills necessary to complete this problem set.

1 Exponential Module

The scalar exponential module maps a scalar variable x to a scalar variable y using the following formula:

$$y = \exp(-\beta x)$$

where β is a parameter.

1.1 Question: jacobian with respect to X

Assuming that the partial derivatives of the energy with respect to y is known, give the expression for the partial derivative of the energy with respect to x. To do so, calculate $\frac{\partial y}{\partial x}$, and apply Chain Rule:

$$\frac{\partial E}{\partial x} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial x}$$

1.2 Question: jacobian with respect to β

Now calculate the derivative of the energy with respect to β:

$$\frac{\partial E}{\partial \beta} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial \beta}$$

2 Component Scaling Module

The scaling module maps an N-dimensional column vector $X = [x_1, x_2, \ldots, x_n]'$ to an N-dimensional column vector $Y = [y_1, y_2, \ldots, y_n]'$ with the following formula:

$$y_i = k_i x_i \quad \forall i \in [1, N].$$

where the k_i's are the components of a vector of parameters $K = [k_1, k_2, \ldots, k_n]$.

2.1 Question: jacobian with respect to X

Assuming that the partial derivatives of the energy with respect to all the y_i are known, give the expression for the partial derivative of the energy with respect to the x_j. To do so, calculate $\frac{\partial y_i}{\partial x_j}$, and apply Chain Rule:

$$\frac{\partial E}{\partial x_j} = \sum_{i=1}^{N} \frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial x_j}$$

Hint: $\frac{\partial y_i}{\partial x_j}$ is equal to 0 for $j \neq i$. you can view the operation on each component $y_i = k_i x_i$ as a separate module operating on scalar values.

2.2 Question: jacobian with respect to K

Now calculate the derivative of the energy with respect to each of the k_j's:

$$\frac{\partial E}{\partial k_j} = \sum_{i=1}^{N} \frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial k_j}$$

hint: $\frac{\partial y_i}{\partial k_j}$ is equal to 0 for $j \neq i$.

3 Global Scaling Module

The global scaling is similar to the component scaling module, except that it uses a single coefficient to scale all the components of X. It maps an N-dimensional column vector $X = [x_1, x_2, \dots, x_n]'$ to an N-dimensional column vector $Y = [y_1, y_2, \dots, y_n]'$ with the following formula:

$$y_i = k x_i \quad \forall i \in [1, N].$$

where the k is a scalar parameters.

3.1 Question: jacobian with respect to X

Assuming that the partial derivatives of the energy with respect to all the y_i are known, give the expression for the partial derivative of the energy with respect to the x_j. To do so, calculate $\frac{\partial y_i}{\partial x_j}$, and apply Chain Rule:

$$\frac{\partial E}{\partial x_j} = \sum_{i=1}^{N} \frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial x_j}$$

3.2 Question: jacobian with respect to K

Now calculate the derivative of the energy with respect to k:

$$\frac{\partial E}{\partial k} = \sum_{i=1}^{N} \frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial k}$$

Note the difference with the component scaling module.

4 Softmax Module

The so-called softmax module maps an N-dimensional column vector $X = [x_1, x_2, \dots, x_n]'$ to an N-dimensional column vector $Y = [y_1, y_2, \dots, y_n]'$ with the following formula:

$$y_i = \frac{\exp(-\beta x_i)}{\sum_{j=1}^{N} \exp(-\beta x_j)} \quad \forall i \in [1, N].$$

where β is a parameter. The softmax module can be used to transform a vector of real numbers into something that looks like a probability distribution: a vector of numbers that are all between 0 and 1 and whose sum is equal to 1.

4.1 Question: jacobian with respect to X

Assuming that the partial derivatives of the energy with respect to all the y_i are known, give the expression for the partial derivative of the energy with respect to the x_j. To do so, calculate $\frac{\partial y_i}{\partial x_j}$, and apply Chain Rule:

$$\frac{\partial E}{\partial x_j} = \sum_{i=1}^{N} \frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial x_j}$$

4.2 Question: jacobian with respect to β

Now calculate the derivative of the energy with respect to β:

$$\frac{\partial E}{\partial \beta} = \sum_{i=1}^{N} \frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial \beta}$$

5 Logsum Module

The logsum module maps an N-dimensional vector X to a scalar y with the following formula:

$$y = -\frac{1}{\beta} \log \left(\sum_{j=1}^{N} \exp(-\beta x_i) \right)$$

5.1 Question: jacobian with respect to X

Assuming that the partial derivatives of the energy with respect to the y is known, give the expression for the partial derivative of the energy with respect to the x_j. To do so, calculate $\frac{\partial y}{\partial x_j}$, and apply Chain Rule:

$$\frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y} \frac{\partial y}{\partial x_j}$$

5.2 Question: jacobian with respect to β

Now calculate the derivative of the energy with respect to β:

$$\frac{\partial E}{\partial \beta} = \frac{\partial E}{\partial y} \frac{\partial y}{\partial \beta}$$