Deep Learning for Master Students – week 4 exercise tasks

Alexander Binder

Week 04: Gradients I

Task1: Some derivatives in math

Compute the 1d derivatives for:

$$f(x) = ax^3 + bx^2 + cx + d, \qquad a, b, c, d, x \text{ all scalars}$$
$$f(x) = e^{-0.5x}\sin(2x),\ x \text{ a scalar}$$

Compute the directional derivatives $Df(\cdot)[\cdot]$ for:

$$f(\mathbf{w}) = g(\mathbf{w}\cdot\mathbf{x} + b), \qquad g(z) = \frac{1}{1+e^{-z}},\ \mathbf{w}\in\mathbb{R}^d, \mathbf{x}\in\mathbb{R}^d,\ b\in\mathbb{R}, z\in\mathbb{R}$$
$$f(\mathbf{X}) = \mathbf{X}\mathbf{a}, \qquad \mathbf{X}\in\mathbb{R}^{d\times k}, \mathbf{a}\in\mathbb{R}^{k\times 1},$$
$$f(\mathbf{X}) = \mathbf{A}\mathbf{X}\mathbf{X}^\top\mathbf{B}, \qquad \mathbf{X}\in\mathbb{R}^{d\times n}, \mathbf{A}\in\mathbb{R}^{d\times d}, \mathbf{B}\in\mathbb{R}^{d\times d}$$

For the last function:

- note that it is a bilinear form in the variables, for which $\mathbf{X}$ is plugged in. $f(\mathbf{X},\mathbf{Y})$ is bilinear, if it is linear in argument $\mathbf{X}$ and also linear in argument $\mathbf{Y}$.
- the directional derivative of a bilinear form $f(\mathbf{X},\mathbf{Y})$:

$$Df(\mathbf{X},\mathbf{Y})[\mathbf{H},\mathbf{K}] = f(\mathbf{H},\mathbf{Y}) + f(\mathbf{X},\mathbf{K})$$

`https://math.stackexchange.com/questions/1120430/derivative-of-bilinear-forms`

Task 2 (Coding): Some derivatives in pytorch

Compute the derivatives in Pytorch for:

$$f(x) = ax^6 + bx^4 + d\ln(x),\ \ a,b,d,x \text{ all scalars}, x > 0$$
$$f(x) = e^{-x}\cos(-2x),\ \ x \text{ a scalar}$$
$$f(\mathbf{w}) = g(\mathbf{w}\cdot\mathbf{x}+b),\ \ g(z) = \frac{1}{1+e^{-z}},\ \ \mathbf{w}\in\mathbb{R}^d, \mathbf{x}\in\mathbb{R}^d, z\in\mathbb{R}, d=3$$
$$f(\mathbf{X}) = \mathbf{vXa}, \mathbf{X}\in\mathbb{R}^{d\times k}, \mathbf{a}\in\mathbb{R}^{k\times 1}, \mathbf{v}\in\mathbb{R}^{1\times d}, k=2, d=3$$
$$f(\mathbf{X}) = \mathbf{vAXX}^\top\mathbf{Bz},\ \ \mathbf{X}\in\mathbb{R}^{d\times n}, \mathbf{A}\in\mathbb{R}^{d\times d}, \mathbf{B}\in\mathbb{R}^{d\times d}, \mathbf{v}\in\mathbb{R}^{1\times d}, \mathbf{z}\in\mathbb{R}^{d\times 1}, n=2, d=3$$

and for some chosen value of the variables.
Steps:

- define a function (or class with additional parameters as class members) which computes $f(\mathbf{w})$ or $f(\mathbf{X})$. `torch.dot`, `torch.mm` or `torch.einsum` can be of help.
  - ask in class if it is not clear for you how to define a class with additional parameters as class members which computes $f(\mathbf{w})$ or $f(\mathbf{X})$
- it has some other parameters, such as v, a, A, B, z, x. For these parameters, seed random generators, then draw the tensors from a zero mean normal distribution with unit covariance or another distribution, in order to set them to some values. You can do the random number generation in PyTorch, numpy is not strictly necessary.
- make sure that the output of the class/function is a scalar, call backward, print the gradient

Task 3: Same thing as Task 1, just differently represented

- Compute the directional derivative $Df(\mathbf{X})[\mathbf{H}]$ in direction $\mathbf{H}$ for:

$$f(\mathbf{X}) = \begin{pmatrix} 1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & x_2^3 \\ \sin x_2 & x_1 \end{pmatrix}$$

Hint:

- what is the dimensionality and shape of the input space ? This will define the dimensionality and shape of the directional vector h
- what is the dimensionality and shape of the output space ? You will have as many output dimensions!

Task 4

- Compute the directional derivatives $Df(\cdot)[\cdot]$ for:

$$
\begin{aligned}
f(\mathbf{X}) &= \mathbf{C}\mathbf{X}\mathbf{B}\mathbf{X}^\top\mathbf{A}\mathbf{X}, \qquad & \mathbf{X} \in \mathbb{R}^{d\times d}, \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \in \mathbb{R}^{d\times d} \\
f(\mathbf{X}) &= \mathbf{X}\mathbf{X}\mathbf{A}\mathbf{X}\mathbf{X}, \qquad & \mathbf{X} \in \mathbb{R}^{d\times d}, \mathbf{A} \in \mathbb{R}^{d\times d}
\end{aligned}
$$

Hint: You can either use product rule and bilinearity, or, deduce a similar approach for 3- and 4-linear mappings.

Task 5 – a dessert

Consider this function from the lecture

$$
f(x,y) = \begin{cases} \frac{y^3}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}
$$

- compute $\delta_{(1,0)}f(0,0) = \frac{\partial f}{\partial x}(0,0)$. This is reallyyyyyy easy, just use the lim-based definition of directional derivative, close your eyes, and receive the present from Nikolaus on your paper (or almost like that).
- compute $\delta_{(0,1)}f(0,0) = \frac{\partial f}{\partial y}(0,0)$. This is easy again, just use the lim-based definition of directional derivative. It is again a straightforward calculation.
- show that the directional derivatives in $x = (0,0)$ do not satisfy a linear relationship. How ?
  - you got already $\delta_{(1,0)}f(0,0)$ and $\delta_{(0,1)}f(0,0)$.
  - take a simple vector $v = (a,b)$ and compute its directional derivative, using the lim-based definition
  - ... and compare that result against the corresponding linear combination result based on the linear combination of $\delta_{(1,0)}f(0,0)$ and $\delta_{(0,1)}f(0,0)$ which reconstructs v.

Task 6 – Coding Bonus

Figure out how to compute a second derivative in Pytorch.
For simplicity you can consider $f(x) = x^\top Bx\|x\|_2^2$ with x being a vector and B of suitable shape so that the result is a scalar.