CS5242 : Neural Networks and Deep Learning

Optional Lecture : Calculus Review

Semester 1 2021/22

Xavier Bresson

https://twitter.com/xbresson

Department of Computer Science
National University of Singapore (NUS)

Outline

- **Partial derivatives**
- **Gradient**
- **Direction of steepest ascent**
- **Chain rule**

Outline

- **Partial derivatives**
- Gradient
- Direction of steepest ascent
- Chain rule

Partial derivatives

$$f(x, y, z) \quad = \quad x^3 y^2 \quad + \quad 2xy^3 z^5$$

$$\frac{\partial f}{\partial y}(x, y, z) \quad = \quad ?$$

Partial derivatives

- **Freeze the x and z variable (view them as constant), then differentiate with respect to the y variable:**

$$f(x,y,z) = x^3y^2 + 2xy^3z^5$$

$$\frac{\partial f}{\partial y}(x,y,z) = x^3\,{\color{blue}2y} + 2x\,{\color{blue}3y^2}z^5$$

$$= 2x^3y + 6xy^2z^5$$

Outline

- Partial derivatives
- **Gradient**
- Direction of steepest ascent
- Chain rule

Gradient

$$f(x, y, z) \;=\; xy^2 + 3x^2 z$$

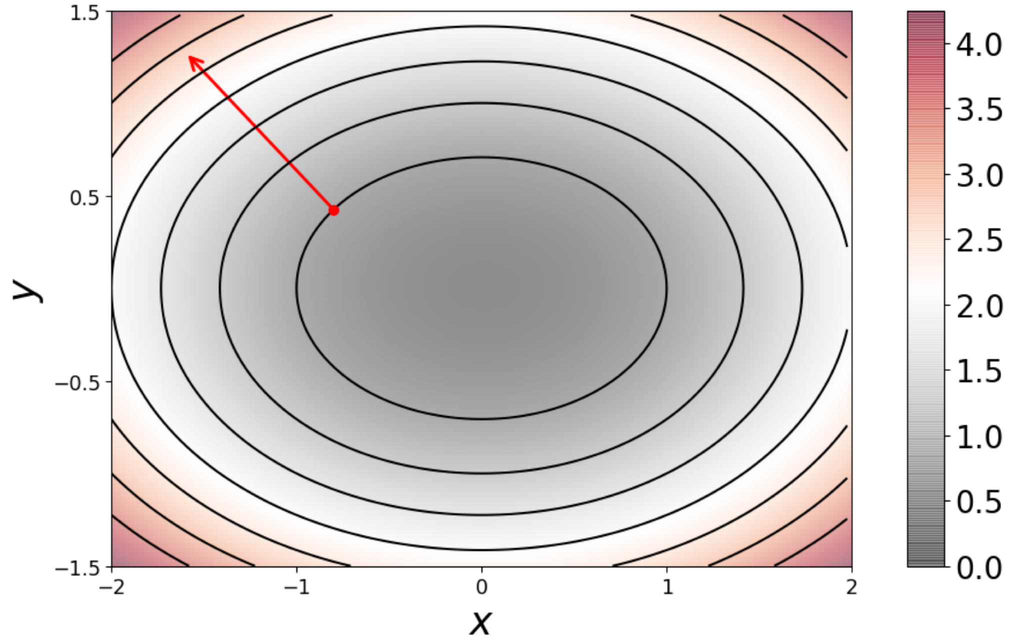
$$\nabla f(x, y, z) \;=\; \begin{bmatrix} \frac{\partial f}{\partial x} \\ \\ \frac{\partial f}{\partial y} \\ \\ \frac{\partial f}{\partial z} \end{bmatrix} \;=\; \begin{bmatrix} y^2 + 6xz \\ \\ 2xy \\ \\ 3x^2 \end{bmatrix} \qquad \nabla f(1, 0, 2) \;=\; \begin{bmatrix} 12 \\ \\ 0 \\ \\ 3 \end{bmatrix}$$

Outline

- Partial derivatives
- Gradient
- **Direction of steepest ascent**
- Chain rule

Direction of steepest ascent

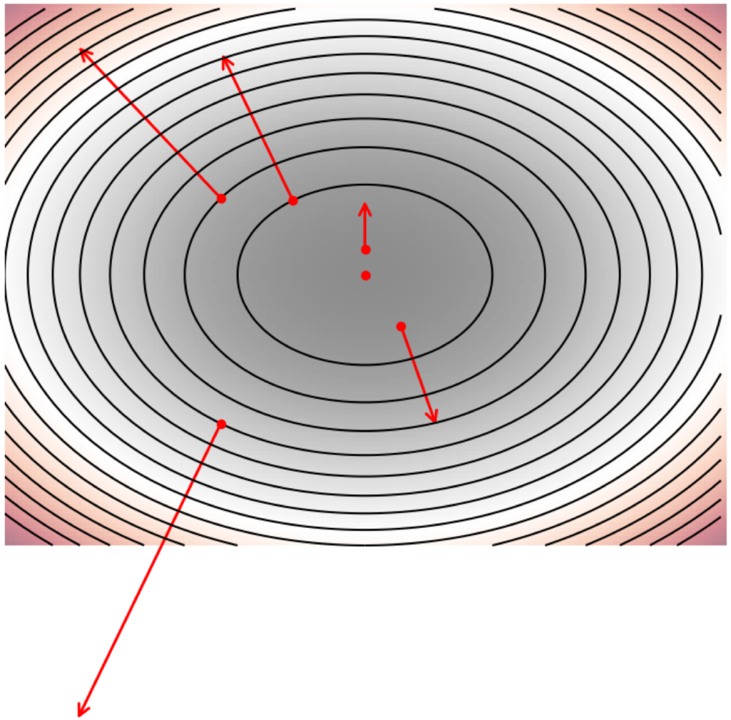
$$f(x,y) = \frac{1}{2}x^2 + y^2 \qquad \nabla f(x,y) = \begin{bmatrix} x \\ 2y \end{bmatrix} \qquad \nabla f(\,-0.8\,,\,0.4\,) = \begin{bmatrix} -0.8 \\ 0.8 \end{bmatrix}$$



Direction of steepest ascent

$$f(x,y) = \frac{1}{2}x^2 + y^2$$

$$\nabla f(x,y) = \begin{bmatrix} x \\ 2y \end{bmatrix}$$



- Gradients are perpendicular to level sets.
- They point in the direction of steepest ascent.
- The steepest is the landscape, the largest is the gradient.
- At the minimum, the gradient is equal to zero.

Outline

- Partial derivatives
- Gradient
- Direction of steepest ascent
- **Chain rule**

Chain rule

$$\frac{d}{dx}\Big[\ f(\ g(x)\ \ )\Big] \ = \ f'(\ g(x)\ )\ \cdot\ g'(x)$$

Example 1:

$$\frac{d}{dx}\Big[\ \ \log(x)\ \ \Big] \ = \ \frac{1}{x}$$

$$\frac{d}{dx}\Big[\ \ \log(1+3x+5x^2)\ \ \Big] \ = \ \frac{1}{1+3x+5x^2}\ \cdot\ (3+10x)$$

Chain rule

- **Example 2:**

$$\frac{d}{dx}\Big[\ \exp(x)\ \Big] \ = \ \exp(x)$$

$$\frac{d}{dx}\Big[\ \exp(ax)\ \Big] \ = \ \exp(ax)\ \cdot\ a$$

Which is better written as:

$$\frac{d}{dx}\Big[\ e^{ax}\ \Big] = e^{ax}\cdot a$$

Chain rule

- **Example 3:**

$$f(x,y,z) = \log(e^{ax} + e^{by} + e^{cz})$$

$$\frac{\partial f}{\partial x}(x,y,z) = \frac{1}{e^{ax} + e^{by} + e^{cz}} \cdot e^{ax} \cdot a$$

$$\frac{\partial f}{\partial y}(x,y,z) = \frac{1}{e^{ax} + e^{by} + e^{cz}} \cdot e^{by} \cdot b$$

$$\frac{\partial f}{\partial z}(x,y,z) = \frac{1}{e^{ax} + e^{by} + e^{cz}} \cdot e^{cz} \cdot c$$

Chain rule

- **Example 4:**

$$L(w_1, w_2, w_3) = \log(e^{w_1 x_1} + e^{w_2 x_2} + e^{w_3 x_3})$$

$$\frac{\partial L}{\partial w_2}(w_1, w_2, w_3) \quad = \quad \frac{1}{e^{w_1 x_1} + e^{w_2 x_2} + e^{w_3 x_3}} \cdot e^{w_2 x_2} \cdot x_2$$

Questions?