GRADIENT

Gradient:

$$f(x) = g_2(g_1(x))$$

Basically calculating partial derivative of nested function

- With vector as input: $f: \mathbb{R}^n \to \mathbb{R}$

$$\nabla_x f(x) = \frac{\partial f(x)}{\partial x}$$

$$= \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \frac{\partial f(x)}{\partial x_2} & \cdots & \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

- With many vectors as input: $f: \mathbb{R}^n \to \mathbb{R}^m$

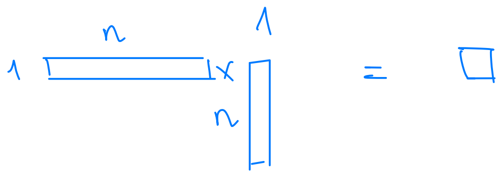
  This is called <u>Jacobian</u>

$$J_f = \nabla_x f(x) = \begin{bmatrix} \nabla_x f_1(x) \\ \nabla_x f_2(x) \\ \vdots \\ \nabla_x f_m(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \ddots & \\ \frac{\partial f_m(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}$$

Size of Gradients (based on input and output size)

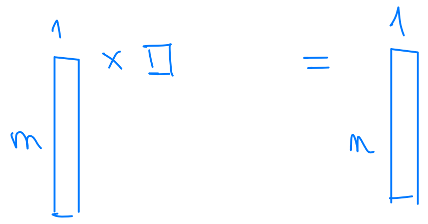
- Gradients of $f: \mathbb{R}^n \to \mathbb{R}$ is size-n row vectors

$$\nabla_x f(x) = [\; \cdots \;]$$

Visual: 

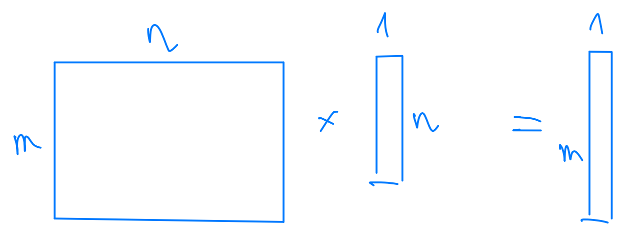
- Partial derivatives of $f: \mathbb{R} \to \mathbb{R}^m$ is size-m column vectors

$$\frac{\partial}{\partial x} f(x) = \begin{bmatrix} \circ \\ \circ \\ \circ \end{bmatrix}$$

Visual: 

- Jacobians of functions $f: \mathbb{R}^n \to \mathbb{R}^m$ is $m \times n$ matrices

$$J_f = \begin{bmatrix} \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \end{bmatrix}$$

Visual: 

Chain Rule:

Given $f(x) = g_2(g_1(x))$ where $\begin{cases} g_1 : \mathbb{R}^n \to \mathbb{R}^m \\ g_2 : \mathbb{R}^m \to \mathbb{R}^k \end{cases}$

The Jacobian of $f(x)$ is:

$$\nabla_x \, g_2(g_1(x))$$
$$= \underbrace{\nabla_y \, g_2(y)}_{J_{g_2} \in \mathbb{R}^{k \times m}} \; \underbrace{\nabla_x \, g_1(x)}_{J_{g_1} \in \mathbb{R}^{m \times n}} \quad \text{where} \quad y = g_1(x)$$