

Math Review

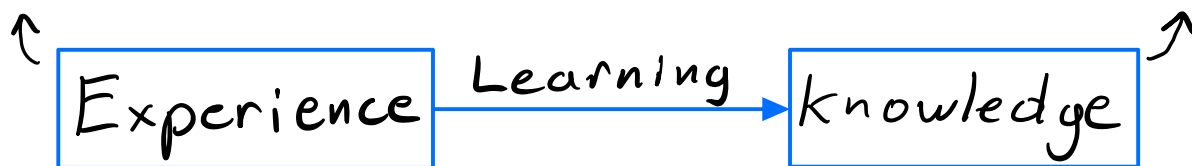
- Sets, tuples, vectors, functions, sums/integrals, derivatives

Motivation

Supervised Learning: Learning from a randomly sampled batch of labeled data

# of rooms	price
2	200
4	590
3	350
7	970

Predictor function f
input: # of rooms
output: price



formalize

Programs/Algorithms
implement functions

Dataset
(tuple of tuples)

Learner
(function)

Predictor
= Model
= Hypothesis
(function)

# of rooms	price
2	200
4	590
3	350
7	970



Sets

Set: a collection of distinct and unordered objects

Ex: $\{0, 1, 2\} = \{0, 1, 2, 2\}$, $\{\text{cat}, \text{dog}\} = \{\text{dog}, \text{cat}\}$

$\mathbb{N} = \{0, 1, 2, \dots\}$ natural numbers

$\mathbb{R}$ real numbers, $\emptyset$ empty set

Variables as sets: $\Omega, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$

Ex: $\mathcal{X} = \{0, 1, 2\}$, $\Omega = \{\text{cat}, \text{dog}\}$

Cardinality: size of the set

Ex: $|\mathcal{X}| = 3$, $|\Omega| = 2$, $|\emptyset| = 0$, $|\mathbb{N}|$, $|\mathbb{R}|$ Fun fact
 $|\mathbb{N}| \neq |\mathbb{R}|$

Countably Infinite Set: If you can list all the elements in the set

Ex: $|\mathbb{N}| = \text{countably infinite}$, $\mathbb{N} = \{0, 1, 2, 3, \dots\}$

$\{x \in \mathbb{N} \mid x \text{ is even}\} = \{0, 2, 4, 6, \dots\}$

Uncountably Infinite Set: Not countably infinite

Ex: $|\mathbb{R}| = \text{uncountably infinite}$, $\mathbb{R} \neq \{0, 0.0001, 0.0002, \dots\}$

$[0, 900] = \text{uncountably infinite}$

Element of and Subset of: $\in, \notin, \subseteq, \subsetneq, \neq$

Ex: $\text{cat} \in \Omega = \{\text{cat}, \text{dog}\}$

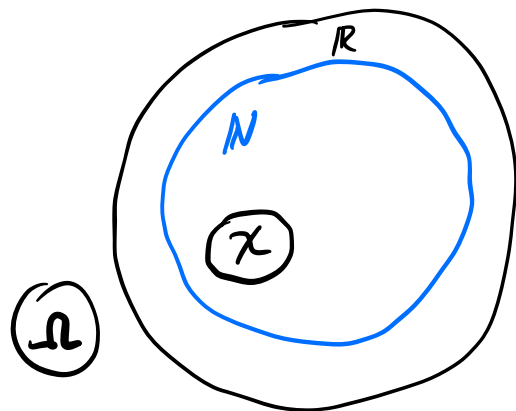
$\text{cat} \notin \mathbb{X} = \{0, 1, 2\}$

$0 \in \mathbb{R}, 1 \in \mathbb{R}, -2 \in \mathbb{R}, \frac{1}{2} \in \mathbb{R}, 0.23 \in \mathbb{R}, \pi \in \mathbb{R}, \infty \notin \mathbb{R}$

$\mathbb{X} \subsetneq \mathbb{N}, \mathbb{X} \subseteq \mathbb{N}, \mathbb{N} \subsetneq \mathbb{R}$

$\mathbb{X} \not\subseteq \Omega, \mathbb{R} \not\subseteq \mathbb{N}$

$\mathbb{X} \subseteq \mathbb{X}, \Omega \subseteq \Omega$



Intervals: Continuous subset of $\mathbb{R}$

Closed: $[0, 1] \subseteq \mathbb{R}$

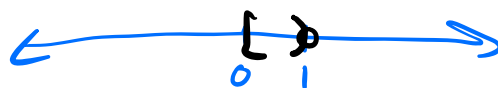


$[0, 1] \not\subseteq \mathbb{N}$

open: $(0, 1) \subseteq \mathbb{R}, 0 \notin (0, 1)$



half-open: $[0, 1) \subseteq \mathbb{R}, 1 \notin [0, 1)$



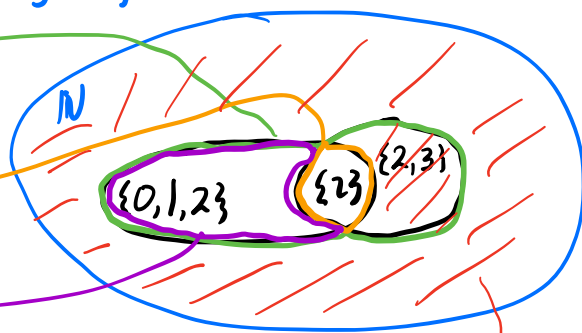
$[0, \infty)$

Unions, Intersections, Set Difference: $\cup, \cap, \setminus$

$\{0, 1, 2\} \cup \{2, 3\} = \{0, 1, 2, 3\}$

$\{0, 1, 2\} \cap \{2, 3\} = \{2\}$

$\{0, 1, 2\} \setminus \{2, 3\} = \{0, 1\}$



Compliment: set^c Need to define a universal set $\mathcal{U}$

Ex: $V=N$, $X^c = \{0,1,2\}^c = V \setminus X = \{3,4,5,\dots\}$ ←

Set Builder Notation: $\{\text{element} \mid \text{property}\}$ ^{"such that"}

Ex: $\{x \in N \mid x \leq 2\} = \{0,1,2\}$

$\{x \in N \mid x \text{ is even}\} = \{0,2,4,\dots\}$

$\{x \in N \mid x \text{ is prime}\} = \{2,3,5,7,11,\dots\}$

$\{x \in N \mid x \in \{0,1,2\} \text{ and } x \notin \{2,3\}\} = \{0,1,2\} \setminus \{2,3\}$

Power set: $P(\mathcal{Z}) = \{S \mid S \subset \mathcal{Z}\}$ the set of all subsets of $\mathcal{Z}$

Ex: $P(\{0,1\}) = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$

Tuple: a collection of ordered objects with duplicates allowed

Ex: $(0,1,2) \neq (0,1,2,2)$, $(\text{cat}, \text{dog}) \neq (\text{dog}, \text{cat})$ This is not an interval!

only variable and element of properties apply

$\Omega' = (0,1,2)$, $X' = (\text{cat}, \text{dog})$, $2 \in \Omega'$, $2 \notin X'$

Invalid: $\Omega' \subset X'$, $\Omega' \cup X'$, $\Omega' \cap X'$, ...

Cartesian products (creating tuples): set $\times$ set

Ex: $\Omega \times X = \{\text{cat}, \text{dog}\} \times \{0,1,2\} \neq X \times \Omega$

$= \{(a,b) \mid a \in \Omega, b \in X\}$

$= \{(a,b) \mid a \in \Omega, b \in \mathcal{Z}\}$

Ex: $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(a, b) \mid a \in \mathbb{R}, b \in \mathbb{R}\}$
 $(0, 2) \in \mathbb{R}^2, (-\frac{1}{10}, \pi) \in \mathbb{R}^2$

Ex: $[0, 1]^2 = [0, 1] \times [0, 1] = \{(a, b) \mid a \in [0, 1], b \in [0, 1]\}$

Ex: $\mathbb{R}^3 = \{(a, b, c) \mid a \in \mathbb{R}, b \in \mathbb{R}, c \in \mathbb{R}\}$

Ex: $\mathcal{X} = \mathbb{R}^2, \mathcal{Y} = \mathbb{R}$

$\mathcal{X} \times \mathcal{Y} = \{(x, y) \mid x \in \mathcal{X}, y \in \mathcal{Y}\} = \mathbb{R}^3$

$(2, 2, 200) = ((2, 2), 200) \in \mathcal{X} \times \mathcal{Y}$

$((2, 12, 1), 200) \notin \mathcal{X} \times \mathcal{Y}$

Ex: $(\mathcal{X} \times \mathcal{Y})^n = (\mathcal{X} \times \mathcal{Y}) \times (\mathcal{X} \times \mathcal{Y}) \times \dots \times (\mathcal{X} \times \mathcal{Y})$ Duplicates are allowed

$D = (((2, 2), 200), ((4, 10), 450), \dots, ((2, 2), 200))$
 $\in (\mathcal{X} \times \mathcal{Y})^n$

Vectors

Motivation: Can model relationships between features and targets

Ex: targets are a linear function of the features (i.e. $y = \vec{x}^T \vec{w}$)

Vector space: A set of tuples that we can add together any elements and multiply any element by a real number (i.e. scalar)

Ex: $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3, \dots$ Why?

Ans($\mathbb{R}^2$): $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2, \vec{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2, c \in \mathbb{R}$

adding: $\vec{x} + \vec{y} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix} \in \mathbb{R}^2$

multiplying by a scalar: $c \vec{x} = c \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} cx_1 \\ cx_2 \end{pmatrix} \in \mathbb{R}^2$

$\Omega = \{\text{dog}, \text{cat}\}$ is not a vector space Why?

Ans: what does $\text{dog} + \text{cat}$ mean?

Vector: An element of a vector space (written as a column)

Ex: $\vec{x} = \begin{pmatrix} 1 \\ 2.5 \end{pmatrix} \in \mathbb{R}^2$, $\vec{w} = \begin{pmatrix} -12 \\ 10 \end{pmatrix} \in \mathbb{R}^2$, $y = 3 \in \mathbb{R}$

Transpose: Changes a column vector to a row vector and vice versa

Ex: $\vec{x}^T = (1, 2.5) \notin \mathbb{R}^2$, $\vec{w}^T = (-12, 10)$, $y^T = 3$

Row vectors belong to a more complicated space so we will not mention it, and instead write the vector space that the column vector belongs to. Ex: $\vec{x} \in \mathbb{R}^2$

Dot Product: A way to multiply two vectors

Ex: $\vec{w}^T \vec{x} = \vec{x}^T \vec{w} = (x_1, x_2) \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = x_1 w_1 + x_2 w_2 = -12 + 25 = 13$

Matrix: Multiple row vectors vertically stacked

Ex: $M = \begin{bmatrix} \vec{x}^T \\ \vec{w}^T \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ w_1 & w_2 \end{bmatrix}$, $M' = \begin{bmatrix} 1 & 2 \\ 1.5 & 3 \\ 0 & -2 \end{bmatrix}$ $\vec{x}^T = (1, 2.5)$

Matrix vector multiplication

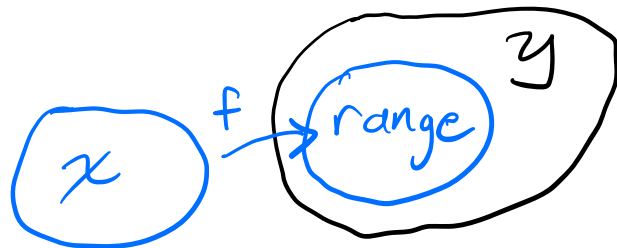
Ex: $M \vec{x} = \begin{bmatrix} \vec{x}^T \\ \vec{w}^T \end{bmatrix} \vec{x} = \begin{pmatrix} \vec{x}^T \vec{x} \\ \vec{w}^T \vec{x} \end{pmatrix}$

$$= \begin{bmatrix} x_1 & x_2 \\ w_1 & w_2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 x_1 + x_2 x_2 \\ w_1 x_1 + w_2 x_2 \end{pmatrix}$$

$$\vec{x} = \begin{pmatrix} 1 \\ 2.5 \end{pmatrix}$$
$$\vec{x} = (1, 2.5)^T$$

Functions

Function: $f: X \rightarrow Y$



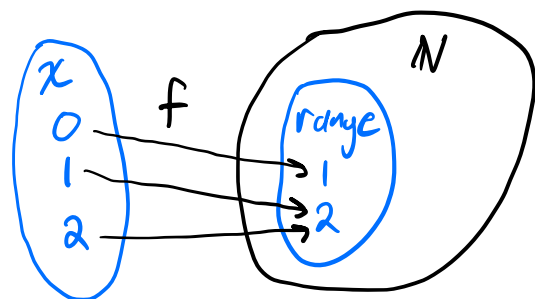
Domain: X Set of all possible inputs to f

Codomain: Y Set of all possible outputs from f

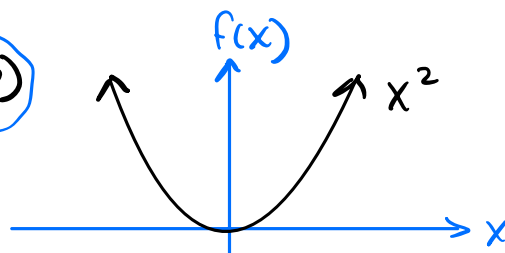
Range: Set of all actual outputs from f

Ex: $f: X \rightarrow Y, X = \{0, 1, 2\}, Y = \mathbb{N}$
 $\text{range} = \{1, 2\}$

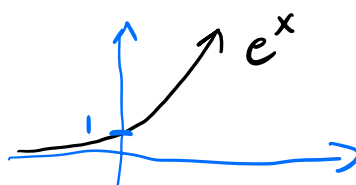
$$f(0) = 1, f(1) = 2, f(2) = 2$$



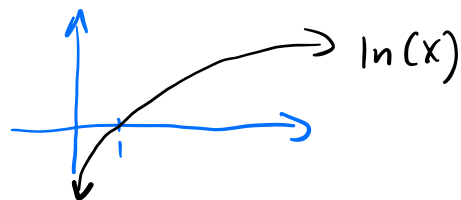
Ex: $f: \mathbb{R} \rightarrow \mathbb{R}$, $\text{range} = \{y \in \mathbb{R} \mid y \geq 0\} = [0, \infty)$
 $f(x) = x^2$ where $x \in \mathbb{R}$



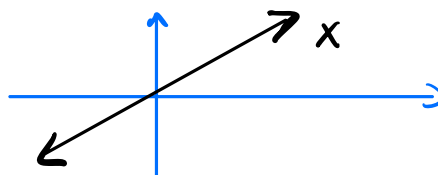
Ex: $f: \mathbb{R} \rightarrow (0, \infty)$, $\text{range} = \{y \in \mathbb{R} \mid y > 0\} = (0, \infty)$
 $f(x) = e^x = \exp(x)$



Ex: $f: X \rightarrow \mathbb{R}, X = (0, \infty), \text{range} = \mathbb{R}$
 $f(x) = \ln(x) = \log_e(x)$

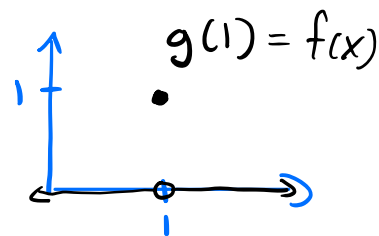


Ex: $f: \mathbb{R} \rightarrow \mathbb{R}, \text{range} = \mathbb{R}$
 $f(x) = \ln(e^x) = x$



Ex: $g: \mathbb{R} \rightarrow \{f \mid f: \mathcal{X} \rightarrow \mathcal{Y}\} \quad \mathcal{X} = \mathbb{R}, \mathcal{Y} = \mathbb{R}$

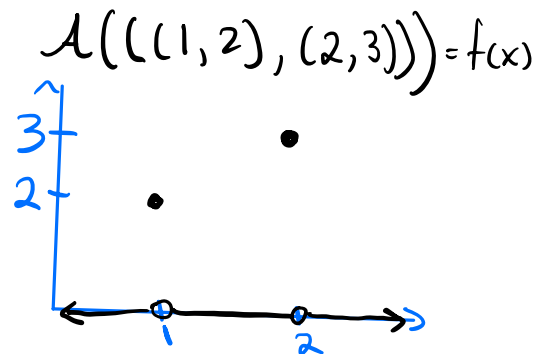
$$g(z) = f \quad \text{where} \quad f(x) = \begin{cases} 1 & \text{if } x = z \\ 0 & \text{otherwise} \end{cases}$$



Ex: $\mathcal{A}: (\mathcal{X} \times \mathcal{Y})^n \rightarrow \{f \mid f: \mathcal{X} \rightarrow \mathcal{Y}\} \quad \mathcal{X} = \mathbb{R}, \mathcal{Y} = \mathbb{R}$

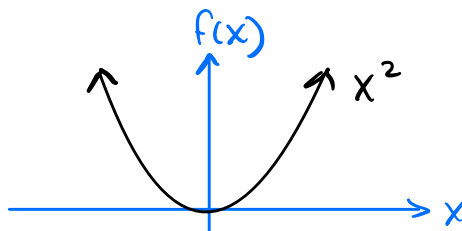
$$\mathcal{A}(D) = \hat{f}, \quad D = ((\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n))$$

$$\hat{f}(x) = \begin{cases} y_i & \text{if } x = x_i \text{ for some } i \in \{1, \dots, n\}. \text{ Pick lowest index} \\ 0 & \text{otherwise} \end{cases}$$

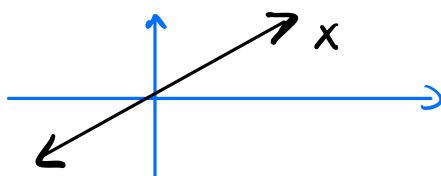


Continuous function: a function without abrupt changes

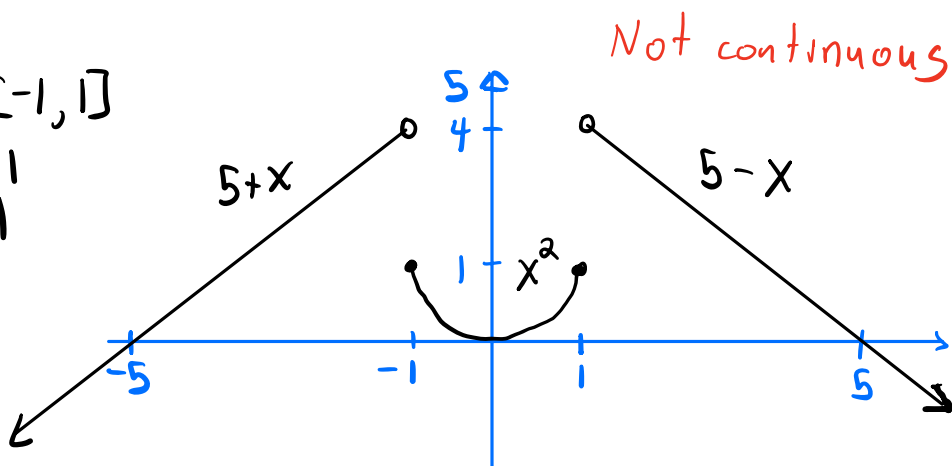
Ex: $f(x) = x^2$



$f(x) = x$



$$f(x) = \begin{cases} x^2 & x \in [-1, 1] \\ 5+x & x < -1 \\ 5-x & x > 1 \end{cases}$$



Summation and Integration: accumulation of values in a set

Motivation: Needed to define expected value

Summation: $\sum$ over discrete sets

Ex: $\mathcal{X} = \{0, 1, 2\}$, $\sum_{x \in \mathcal{X}} x = 0 + 1 + 2 = 3$

$$f(x) = x^2, \sum_{x \in \mathcal{X}} f(x) = 0^2 + 1^2 + 2^2 = 5$$

$$\mathcal{X} = (x_1, x_2, \dots, x_n), \sum_{i=1}^n x_i = \sum_{x \in \mathcal{X}} x = x_1 + x_2 + \dots + x_n$$

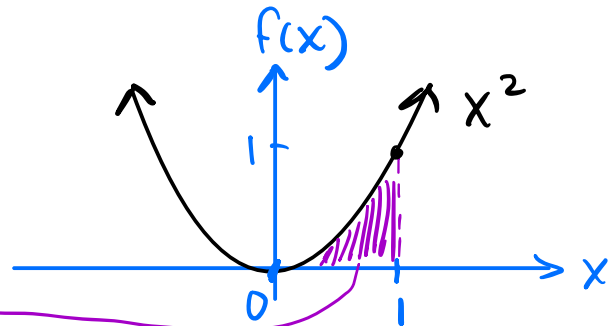
Integration: $\int$ over continuous sets

Ex: $\mathcal{X} = [a, b]$, $f: \mathcal{X} \rightarrow \mathbb{R}$

$$\int_{\mathcal{X}} f(x) dx = \int_a^b f(x) dx$$

if $a=0, b=1, f(x)=x^2$, then

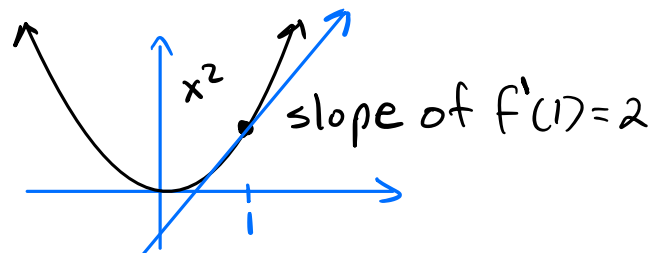
$$\int_0^1 x^2 dx = \left. \frac{x^3}{3} \right|_0^1 = \frac{1}{3}$$



Derivatives: rate of change of a function
written f' or $\frac{df}{dx}$ for a function f

Motivation: We want our learner to pick the best predictor (optimization)

Ex: $f(x) = x^2$, $f'(x) = 2x = \frac{df}{dx}(x)$



$$f(x) = x^a, f'(x) = ax$$

$$f(x) = e^x, f'(x) = e^x$$

$$f(x) = \ln(x), f'(x) = \frac{1}{x}$$

Chain rule $f(x) = g(h(x)), f'(x) = g'(h(x)) h'(x)$

Ex: $f(x) = \exp(x^2), g(h(x)) = \exp(h(x)), h(x) = x^2$

$$g'(h(x)) = \exp(h(x)), h'(x) = 2x$$

$$f'(x) = \exp(x^2) 2x$$

Partial Derivative: Derivative of a function that takes as input more than one variable

$\frac{\partial f}{\partial x_i}(x_1, x_2)$ is the partial derivative of $f(x_1, x_2)$ with respect to x_i

Ex: $f(x_1, x_2) = 2x_1 x_2^2, \frac{\partial f}{\partial x_1}(x_1, x_2) = 2x_2^2, \frac{\partial f}{\partial x_2}(x_1, x_2) = 4x_1 x_2$

Often we write $\frac{\partial f}{\partial x_i}$ instead of $\frac{\partial f}{\partial x_i}(x_1, x_2)$

Ex: $f(\vec{x}, \vec{w}) = f(x_1, \dots, x_d, w_1, \dots, w_d)$

$$= x_1 w_1 + \dots + x_d w_d = \sum_{i=1}^d x_i w_i = \vec{x}^T \vec{w}$$

where $\vec{x} = (x_1, \dots, x_d)^T \in \mathbb{R}^d$ and $\vec{w} = (w_1, \dots, w_d)^T \in \mathbb{R}^d$

$$\frac{\partial f}{\partial w_1} = x_1, \dots, \frac{\partial f}{\partial w_d} = x_d$$

Gradient: A vector of all the partial derivatives

$$\nabla f(\vec{x}) = \nabla f(x_1, \dots, x_d) = \left(\frac{\partial f}{\partial x_1}(x_1), \dots, \frac{\partial f}{\partial x_d}(x_d) \right)^T$$

Ex: $f(\vec{x}) = f(x_1, x_2) = 2x_1x_2^2$

$$\nabla f(\vec{x}) = (2x_2^2, 4x_1x_2)^T$$