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

Experience $\xrightarrow{\text{Learning}}$ Knowledge

formalize

Programs/Algorithms implement functions

Dataset (tuple of tuples) $\xrightarrow{\text{Learner (function)}}$ Predictor = Model = Hypothesis (function)

$$\boxed{\begin{aligned}&D = ((\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)) \\ &D \in (\mathcal{X} \times \mathcal{Y})^n\end{aligned}}$$

D n feature-label pairs

$\mathcal{X}$ set of d-dimensional features

$\mathcal{Y}$ set of labels/targets

features	label/target
$\vec{x}_1 = (x_{1,1}, \ldots, x_{1,d})^T$	y_1
$\vec{x}_2 = (x_{2,1}, \ldots, x_{2,d})^T$	y_2
$\vdots \quad \vdots$	$\vdots$
$\vec{x}_n = (x_{n,1}, \ldots, x_{n,d})^T$	y_n

$\vec{x}_i \in \mathcal{X} = \mathbb{R}^d$, $y_i \in \mathcal{Y} = \mathbb{R}$

$i \in \{1, \ldots, n\}$

$$\boxed{f : \mathcal{X} \to \mathcal{Y}}$$

f: a function from features to labels

Ex: $200x + 100$, $\mathcal{X} = \mathbb{R}$, $\mathcal{Y} - \mathbb{R}$

$$\boxed{\mathcal{A} : (\mathcal{X} \times \mathcal{Y})^n \to \{f \mid f : \mathcal{X} \to \mathcal{Y}\}}$$

$\mathcal{A}$: a function from datasets to predictors

$\mathcal{A}(D) = \hat{f}$ where $\hat{f} : \mathcal{X} \to \mathcal{Y}$

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

Sets

Set: a collection of distinct and unordered objects

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

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

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

Variables as sets: Ω, X, Y, Z

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

Cardinality: size of the set

countably infinite

uncountably infinite

Ex: $|X| = 3$, $|\Omega| = 2$, $|\phi| = 0$, $|\mathbb{N}|$, $|\mathbb{R}|$

Fun fact

$|\mathbb{N}| \neq |\mathbb{R}|$

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

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

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

Uncountably Infinite Set: Not countably infinite

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

$|[0,900]|$ = uncountably infinite

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

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

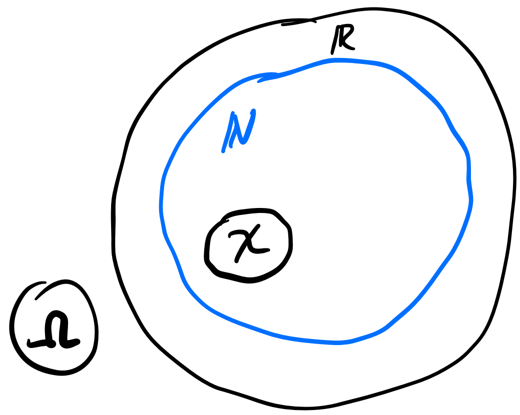
$cat \notin \mathcal{X} = \{0,1,2\}$

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

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

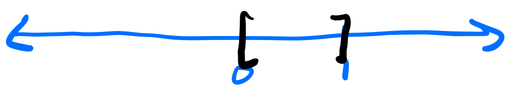
$\mathcal{X} \nsubseteq \Omega, \mathbb{R} \nsubseteq \mathbb{N}$

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



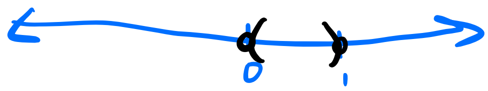
Intervals: Continuous subset of $\mathbb{R}$

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

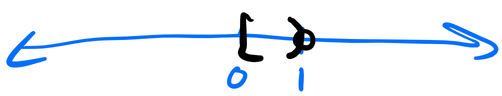


$[0,1] \nsubseteq \mathbb{N}$

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



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



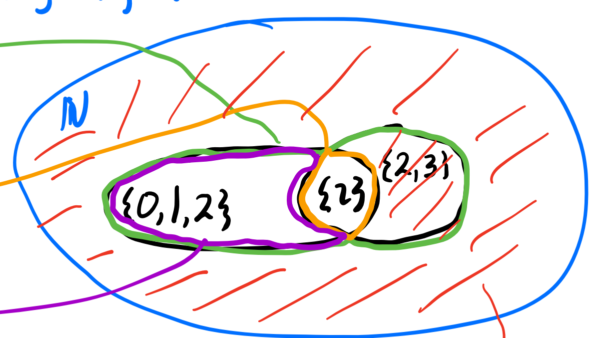
$[0,\infty)$

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: $\mathcal{U}=\mathbb{N}$, $\mathcal{X}^c=\{0,1,2\}^c=\mathcal{U}\setminus\mathcal{X}=\{3,4,5,\ldots\}$

Set Builder Notation: {element | property}   "such that"

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

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

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

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

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

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

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

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

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

Cartesian products (creating sets of tuples): set × set

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

$=\{(\text{cat},0),(\text{cat},1),(\text{cat},2),(\text{dog},0),(\text{dog},1),(\text{dog},2)\}$

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

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

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

$\mathcal{X}\times\mathcal{Y} = \{(x,y) \mid x\in\mathcal{X}, y\in\mathcal{Y}\} = \mathbb{R}^3 = \mathbb{R}^2\times\mathbb{R} = (\mathbb{R}\times\mathbb{R})\times\mathbb{R}$

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

$\mathbb{R}^4 \ni ((2,12,1),200) \notin \mathcal{X}\times\mathcal{Y}$

Ex: $(\mathcal{X}\times\mathcal{Y})^n = \underset{1}{(\mathcal{X}\times\mathcal{Y})}\times\underset{2}{(\mathcal{X}\times\mathcal{Y})}\times\underset{\cdots}{\cdots}\times\underset{n}{(\mathcal{X}\times\mathcal{Y})}$ Duplicates are allowed

$D = \Big(\big(\underset{=\vec{x}_1}{(2,2)},\underset{=y_1}{200}\big), \big(\underset{=\vec{x}_2}{(4,10)},\underset{=y_2}{450}\big), \ldots, \big(\underset{=\vec{x}_n}{(2,2)},\underset{=y_n}{200}\big)\Big) \in (\mathcal{X}\times\mathcal{Y})^n$

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: $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3, \ldots$ 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: $c\vec{x} = c\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} cx_1 \\ cx_2 \end{pmatrix} \in\mathbb{R}^2$

$\Omega = \{dog, cat\}$ is not a vector space why?
Ans: what does dog + 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: $\underset{\vec{w}\cdot\vec{x}}{\underset{\shortparallel}{\vec{w}^T\vec{x}}} = \vec{x}^T\vec{w} = (x_1, x_2)\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = x_1 w_1 + x_2 w_2 = -12 + 25 = 13$

$(1, 0)\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$ $\quad$ xw

Matrix: Multiple row vectors vertically stacked

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

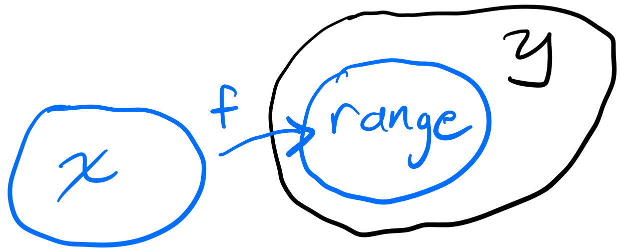
Matrix vector multiplication

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

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

Functions

$f(x)=1$



Function: $f: \underset{``\mathbb{R}}{\mathcal{X}} \to \underset{``\mathbb{R}}{\mathcal{Y}}$

Domain: $\mathcal{X}$ Set of all possible inputs to f

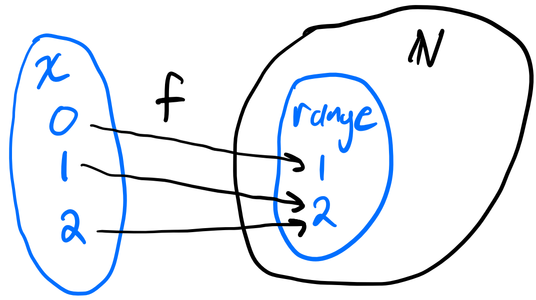
Codomain: $\mathcal{Y}$ Set of all possible outputs from f

Range: Set of all actual outputs from f

Ex: $f: \mathcal{X} \to \mathcal{Y}$, $\mathcal{X} = \{0,1,2\}$, $\mathcal{Y} = \mathbb{N}$,

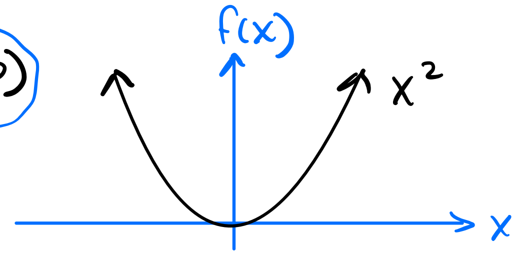
$\text{range} = \{1,2\}$

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



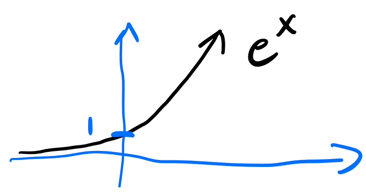
Ex: $f: \mathbb{R} \to \mathbb{R}$, $\text{range} = \{y \in \mathbb{R} \mid y \geq 0\} = [0,\infty)$

$f(x) = x^2$ where $x \in \mathbb{R}$



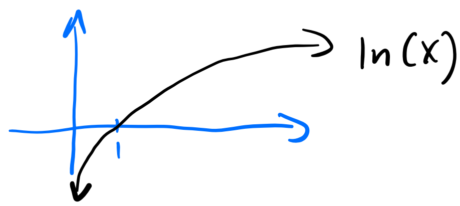
Ex: $f: \mathbb{R} \to (0,\infty)$ $\text{range} = \{y \in \mathbb{R} \mid y > 0\} = (0,\infty)$

$f(x) = e^x = \exp(x)$



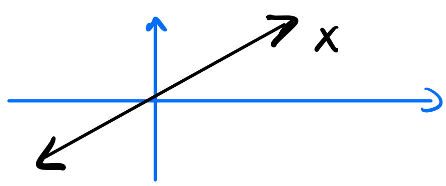
Ex: $f: \mathcal{X} \to \mathbb{R}$, $\mathcal{X} = (0,\infty)$, $\text{range} = \mathbb{R}$

$f(x) = \ln(x) = \log_e(x)$



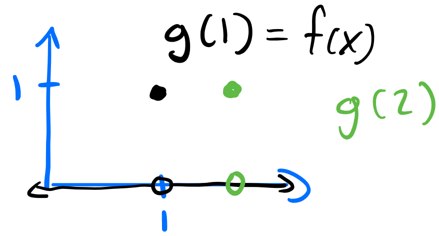
Ex: $f: \mathbb{R} \to \mathbb{R}$, $\text{range} = \mathbb{R}$

$f(x) = \ln(e^x) = x$



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

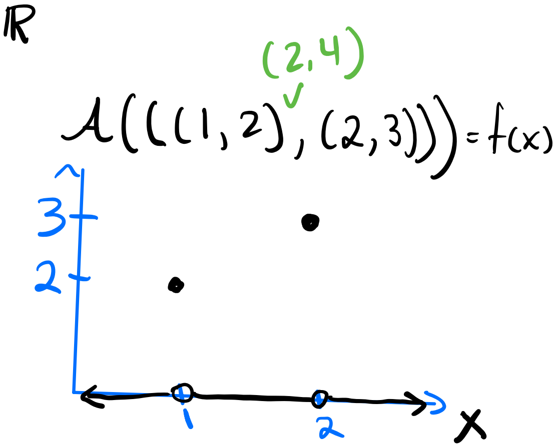
$$g(z) = f \quad \text{where} \quad f(x) = \begin{cases} 1 & \text{if } x = z \\ 0 & \text{otherwise} \end{cases}$$



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

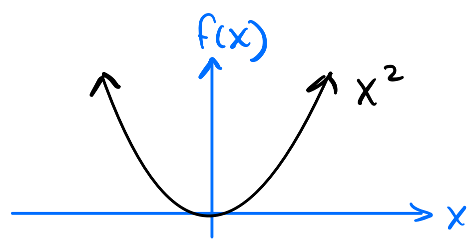
$$\mathcal{A}(D) = \hat{f}, \quad D = ((\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n))$$

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

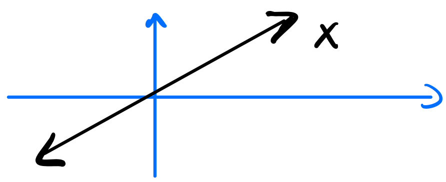


Continuous function: a function without abrupt changes

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

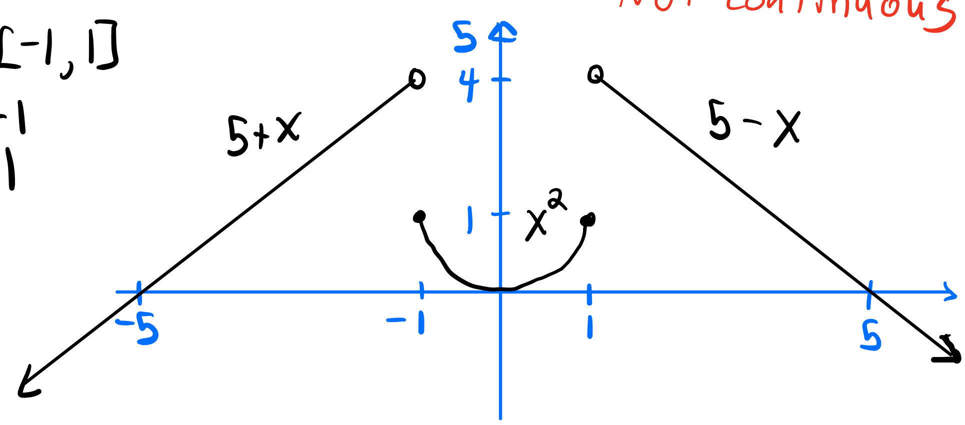


$f(x) = x$



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

Not continuous



Summation and Integration: accumulation of values in a set

Motivation: Needed to define expected value

Summation: Σ over discrete sets

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

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

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

$\cancel{\sum_{x \in \mathbb{R}}}$

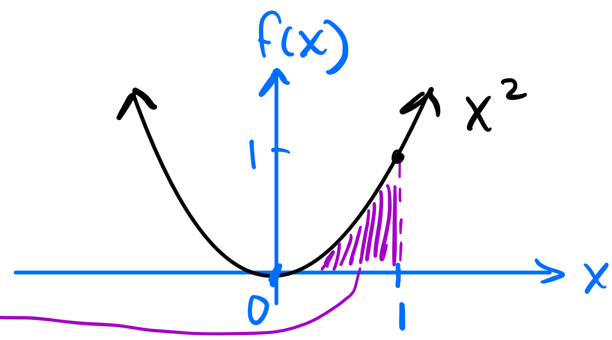
Integration: $\int$ over continuous sets

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

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

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

$\int_0^1 x^2\,dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$

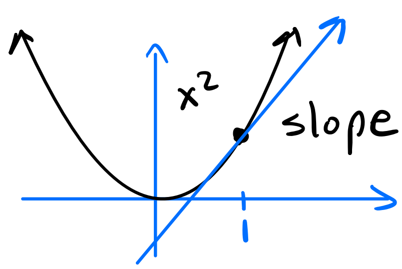


Derivatives: rate of change of a function

written f' or $\frac{df}{dx}$ for a function f

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

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



slope of $f'(1) = 2$

$f(x) = x^a, \; f'(x) = a x^{a-1}$

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

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

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

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

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

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

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

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

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

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

Gradient: A vector of all the partial derivatives

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

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

$$\nabla f(\vec{x}) = \left(2x_2^2, 4x_1 x_2\right)^T$$