Important Announcements

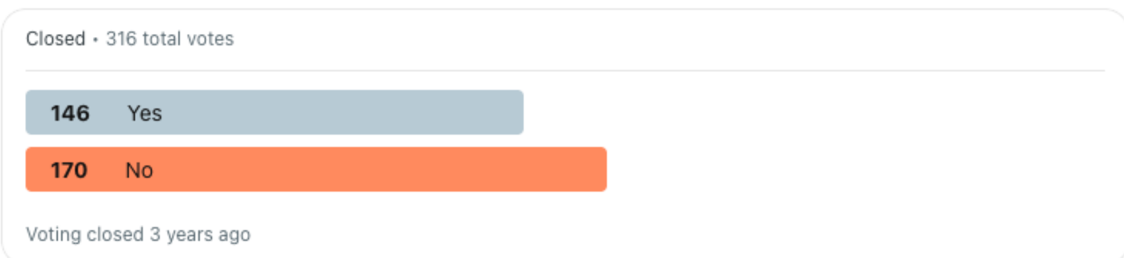
- assignment 1 will be released this Friday (Jan 17)
- videos & slides/notes are online

regarding last lecture

- input to learner function
- 2 is also a prime number

$\mathbb{N}^0$

Do you usually include 0 in the set of natural numbers?



- brackets for sets, intervals, tuples
- why are duplicates important when defining a dataset?
- what is a universal set?
- duplicates & order
- difference between countably infinite & uncountably infinite sets?
- $\mathcal{X} \times \mathcal{Y} = (\mathbb{R}^2) \times \mathbb{R} \neq \mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3$, $(2,2,200) \in \mathbb{R}^3$

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

$\mathcal{X} \times \mathcal{Y} = \{(x,y) \mid x \in \mathcal{X}, y \in \mathcal{Y}\}$ our dataset from before

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

$\mathcal{X} \times \mathcal{Y} = (\mathbb{R}^2) \times \mathbb{R} \neq \mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3$, $(2,2,200) \in \mathbb{R}^3$

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

$D = (((2,2),200), ((4,10),450), \ldots, ((2,2),200))$

$\mathcal{X} = \{0.1, 0.2, 0.5, 25, \ldots, x_{t+1}+5\}$

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?

$\oplus$ mathematical operations on tuples

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

sometimes called inner product

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

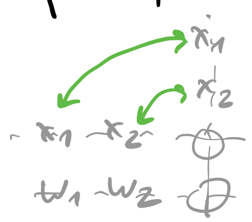
Matrix: Multiple row vectors vertically stacked

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

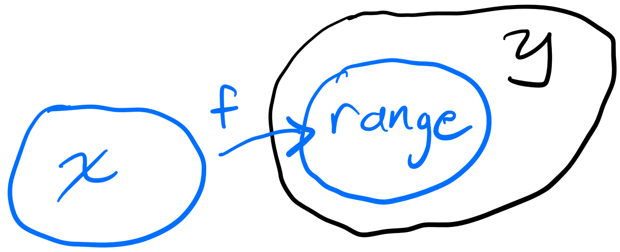
Matrix vector multiplication

Ex: $M\vec{x} = \begin{bmatrix} \vec{x}^T \\ \vec{w}^T \end{bmatrix} \vec{x} = \begin{bmatrix} x_1 & x_2 \\ w_1 & w_2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 x_1 + x_2 x_2 \\ w_1 x_1 + w_2 x_2 \end{pmatrix}$

$$= \begin{pmatrix} \vec{x}^T \vec{x} \\ \vec{w}^T \vec{x} \end{pmatrix}$$



Functions



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

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

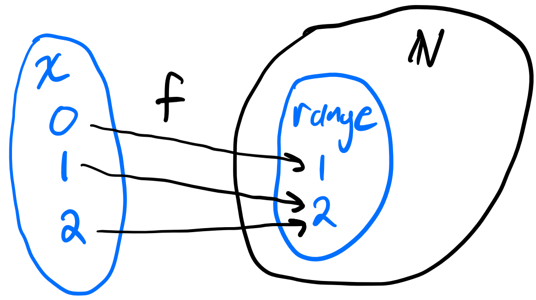
Codomain: $\mathcal{Y}$ Set of all possible outputs from f

Range: Set of all actual outputs from f

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

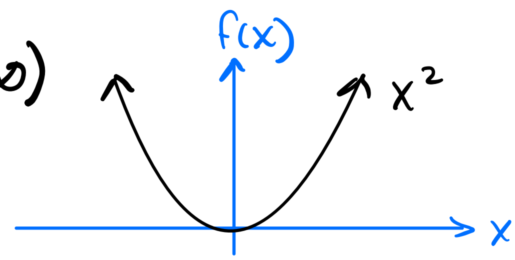
$$\text{range} = \{1, 2\}$$

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



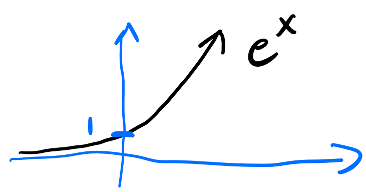
Ex: $f: \mathbb{R} \to \mathbb{R}$, $\text{range} = \{y \in \mathbb{R} \mid y \geq 0\} = [0, \infty)$

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



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

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



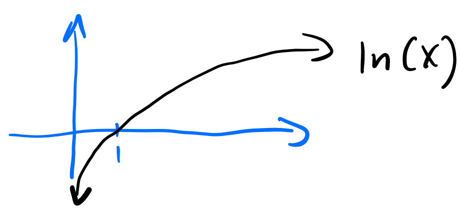
Ex: $f: \mathcal{X} \to \mathbb{R}$, $\mathcal{X} = (0, \infty)$, $\text{range} = \mathbb{R}$

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

$e^a = x$

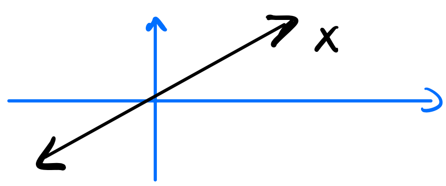
$\Rightarrow a = \log_e(x)$

$= \ln(x)$



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

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

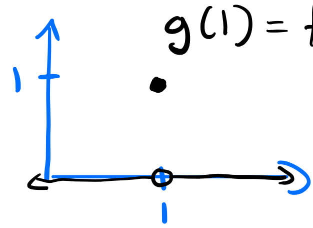


Ex: $g : \mathbb{R} \to \{f \mid f : \mathcal{X} \to \mathcal{Y}\}$ $\mathcal{X} = \mathbb{R}, \mathcal{Y} = \mathbb{R}$

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

learner

$g(1) = f(x)$

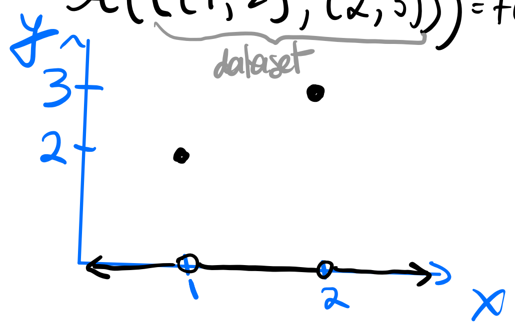


Ex: $\mathcal{A} : (\mathcal{X} \times \mathcal{Y})^n \to \{f \mid f : \mathcal{X} \to \mathcal{Y}\}$ $\mathcal{X} = \mathbb{R}, \mathcal{Y} = \mathbb{R}$

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

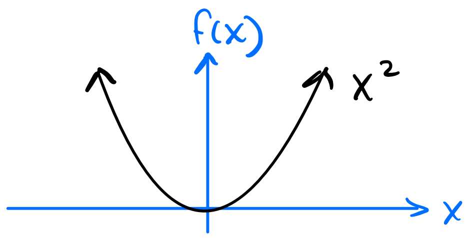
$f(x) = \begin{cases} y_i & \text{if } x = x_i \text{ for some } i \in \{1, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}$

$\mathcal{A}(\underbrace{((1,2),(2,3))}_{\text{dataset}}) = f(x)$

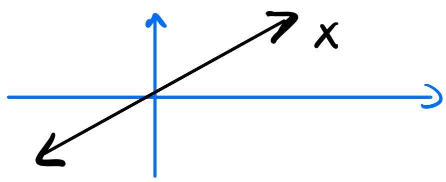


Continuous function: a function without abrupt changes

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

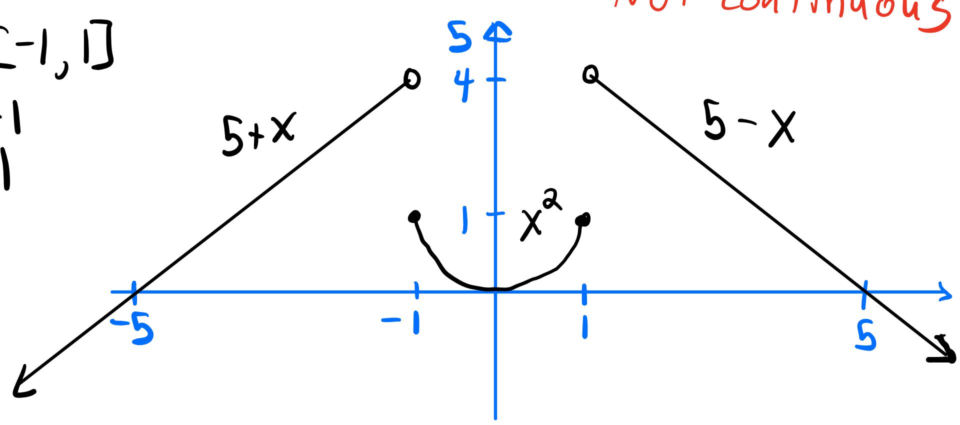


$f(x) = x$



Not continuous

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



Summation and Integration: accumulation of values in a set

Motivation: Needed to define expected value

Summation: $\sum$ over discrete sets finite or countably infinite

important for expected values

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

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

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

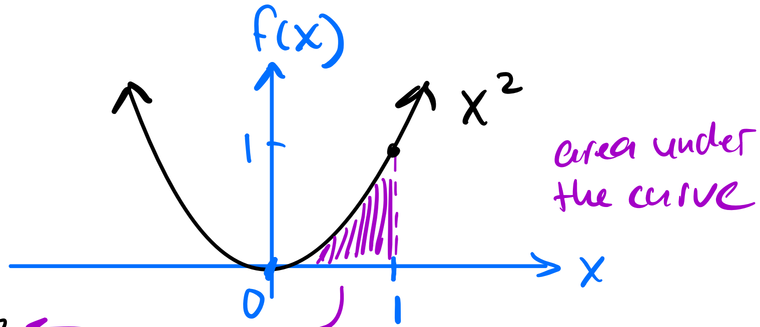
Integration: $\int$ over continuous sets uncountably infinite sets

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

$$\int_0^1 x^2\,dx = \frac{x^3}{3}\Big|_0^1 = \frac{1^3}{3} - \frac{0^3}{3} = 1/3$$

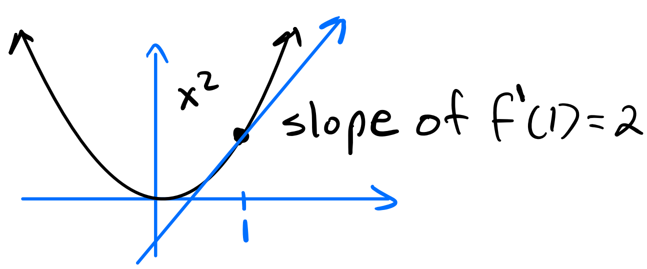


Derivatives: rate of change of a function

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

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

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



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

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

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

$h(x) = u$

Chain rule $f(x) = g(h(x)), f'(x) = g'(h(x))\, h'(x)$ OR $\frac{d\, g(h(x))}{dx} = \frac{dg(u)}{du}\frac{du}{dx}$

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

$$\frac{d}{du} g(u) = \frac{d}{du}\exp(u) = \exp(u) = \exp(x^2), \quad \frac{du}{dx} = h'(x) = 2x$$

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

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: $f(x_1, x_2) = 2x_1 + 3x_2^2, \quad \frac{\partial f}{\partial x_1}(x_1) = 2, \quad \frac{\partial f}{\partial x_2}(x_2) = 6x_2$

Ex: $f(\vec{x}) = f(x_1, \dots, x_d) = x_1 w_1 + \dots + x_d w_d = \sum_{i=1}^{d} x_i w_i = \vec{x}^T \vec{w}$

where $\vec{x} = (x_1, \dots, x_d)^T \in \mathbb{R}^d$ are the variables

and $\vec{w} = (w_1, \dots, w_d)^T \in \mathbb{R}^d$ are some fixed numbers (constants)

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

Gradient: A vector of all the partial derivatives

$$\nabla f(x_1, x_2) = \left(\frac{\partial f}{\partial x_1}(x_1), \frac{\partial f}{\partial x_2}(x_2) \right)^T$$

<u>Ex</u>: $f(\vec{x}) = f(x_1, \dots, x_d) = x_1 w_1 + \dots + x_d w_d = \vec{x}^T \vec{w}$

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

$$= (w_1, \dots, w_d)^T$$