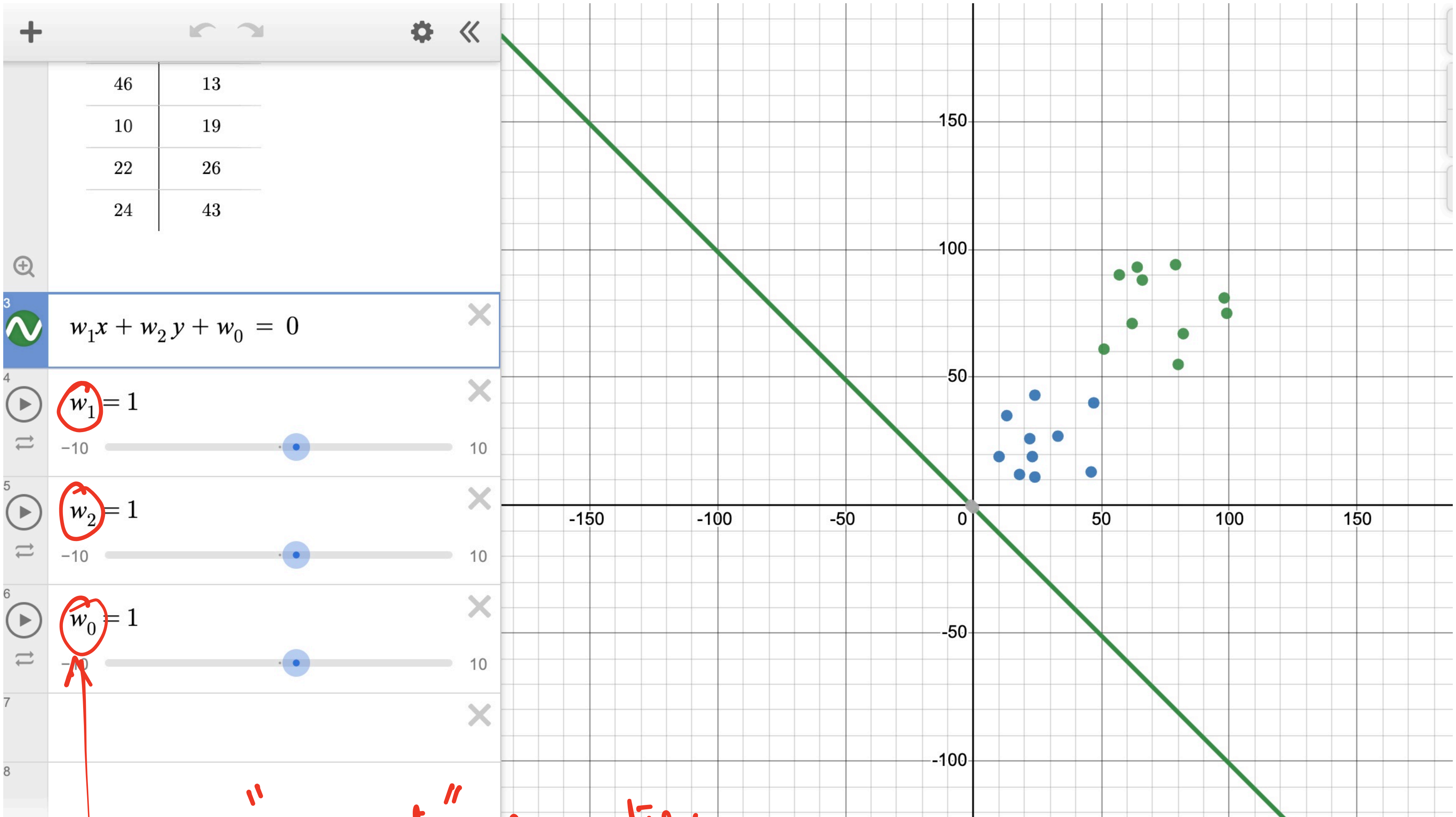
February 28, 2023 DSML : Math for ML.

Optimization 1 : The need for Calculus in ML.

Agenda for today :

(a) Time complexity of Brute force optimization.
(b) Mathematical Definition of a Binary classifiers.
(c) Functions, Limits and Continuity.

Basic intuition for Classification



pick the "correct" parameters so that we can identify the class.

Simplest searching algorithm : linear search.

Q] Can we do optimization using linear search?

If I am searching, what is the data structure?

$w_1 \rightarrow$ 201 values between [-10, 10] where we consider [-10, -9.9, -9.8, ..., 0, ... 9.9, 10]

$w_2 \rightarrow$ 201 values.

$w_0 \rightarrow$ 201 values.

Total possibilities : $(201)^3$

10^{-6} s per possibility.

$[(201)^3 \times 10^{-6}]$

"Curse of dimensionality"

→ It takes roughly years to deal with # parameters > 6.

→ If we do this with just 6 parameters, we will need 763 days to check all possibilities!

How to solve optimization problems?

→ Use Gradient Descent!!

↳ It allows us to search for an approximately correct solution very fast.

maxima, minima.

↳ calculus in multiple variables.

↳ " ——— " ——— single variable.

↳ Derivatives, slopes, tangents.

↓

Limits, continuity, differentiability etc.

Defining the Classification problem mathematically.

Given: Labelled dataset.

$$\mathcal{D} = \{ (\underbrace{\bar{x}_i}_{\text{feature vector}}, \underbrace{y_i}_{\text{label}}) : \bar{x}_i \in \mathbb{R}^d, y_i \in \{-1, 1\} \}_{i=1}^{n}$$

Goal: Find a function $f(\bar{x})$ such that

classifier. $\leftarrow$ $\boxed{f(\bar{x}_i)} = y_i$ ← actual label.

$sign(\bar{w}^T x_i + w_0) \rightarrow$ ideally be y_i

$y_i' \rightarrow$ predicted label.

this function measures how well classifier is doing.

$g(\mathcal{D}; \bar{w}, w_0) \rightarrow$ gain function, this is the function we optimize.

$$g(\bar{x}_i, y_i; w_0, \bar{w}) = \left(\frac{\bar{w}^T\bar{x}_i + w_0}{\|\bar{w}\|} \cdot y_i\right)$$

$w_0^*, \bar{w}^* \rightarrow$ Optimal answers for "Best Classifier"

$$\boxed{w_0^*, \bar{w}^* = \underset{\bar{w}, w_0}{\arg\max}\ g(D; \bar{w}, w_0)}$$

The best possible classifier is defined by $w_0^*, \bar{w}^*$ which gives the maximum value for the gain function.

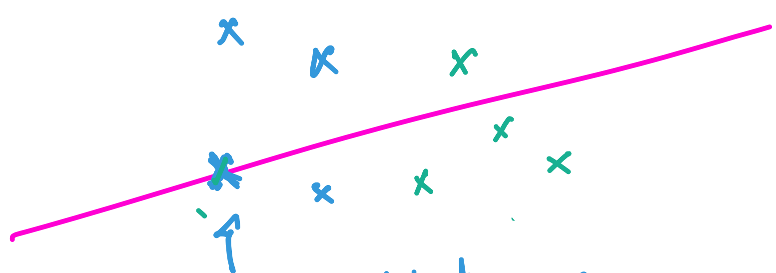
What is the relationship between distance & gain function?

The gain function we will optimize.

① $$\ell(D; w_0, \bar{w}) = \sum_{i=1}^{n} g(\bar{x}_i, y_i; w_0, \bar{w})$$

② $$\ell(D; w_0, \bar{w}) = \prod_{i=1}^{n} \underbrace{g(x_i, y_i; w_0, \bar{w})}_{X}$$

0.



then distance is 0. So, we reject option 2.

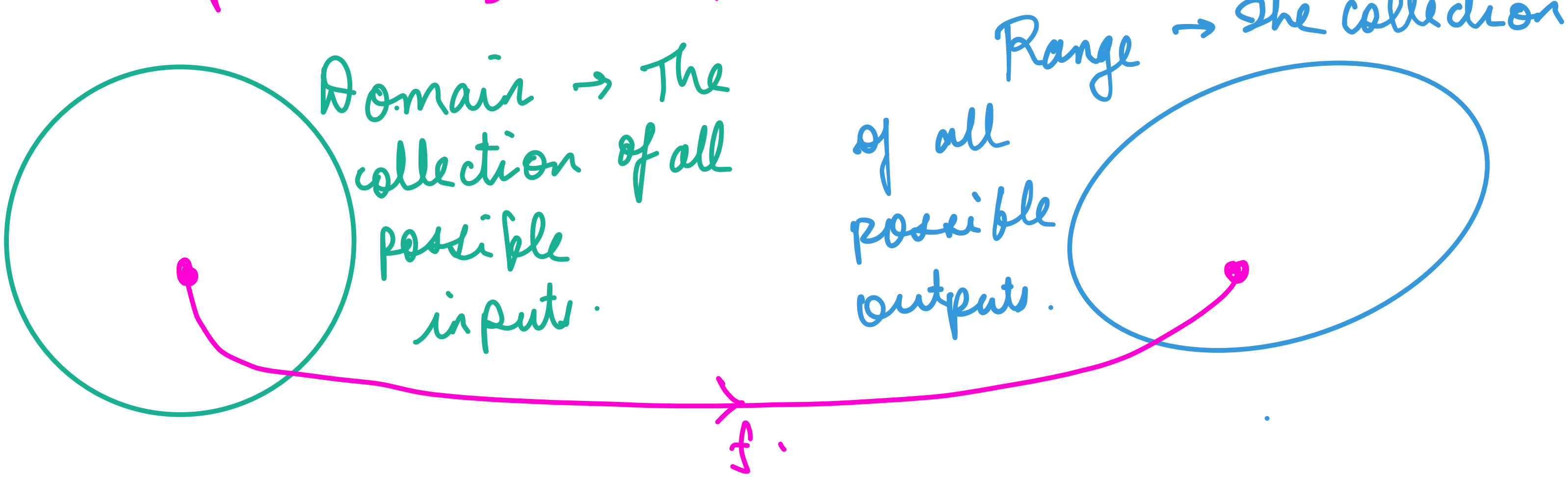
Functions:

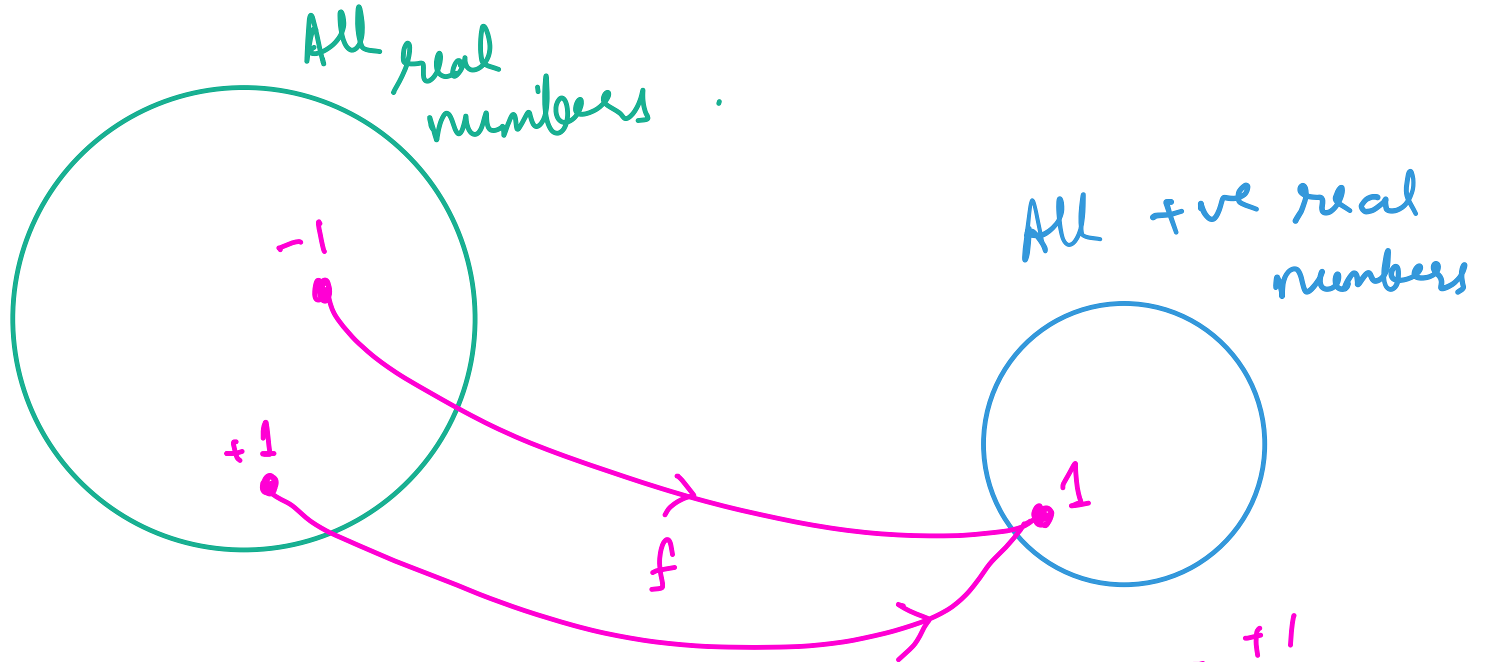
$$f(x) = y \leftarrow \text{output.}$$

$\uparrow$

Input

"A function is a mapping between input & output."

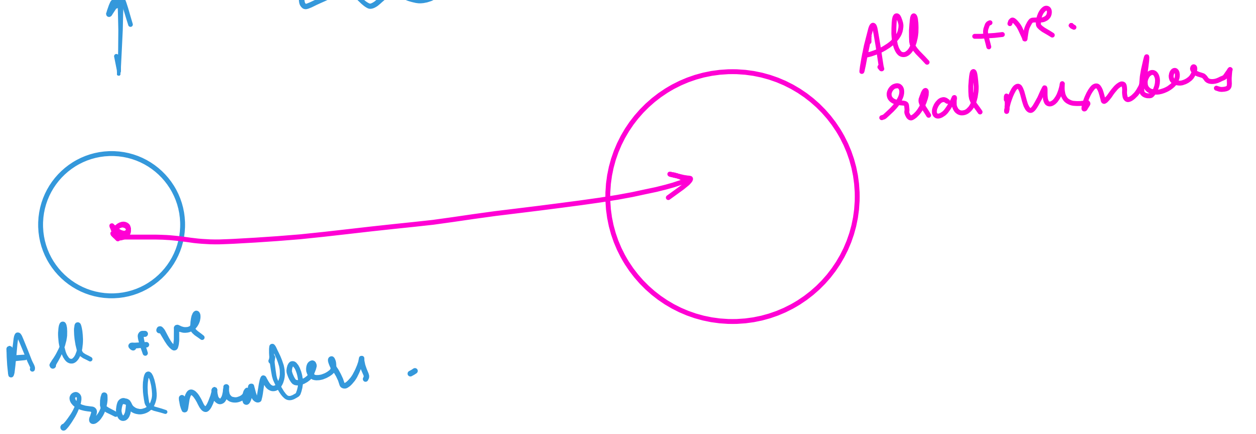




$1 \to +1$

$f(x) = x^2$

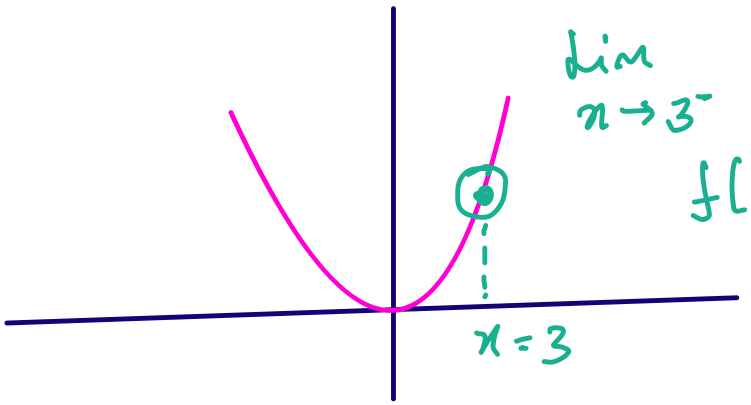
$f(x) = +\sqrt{x}.$



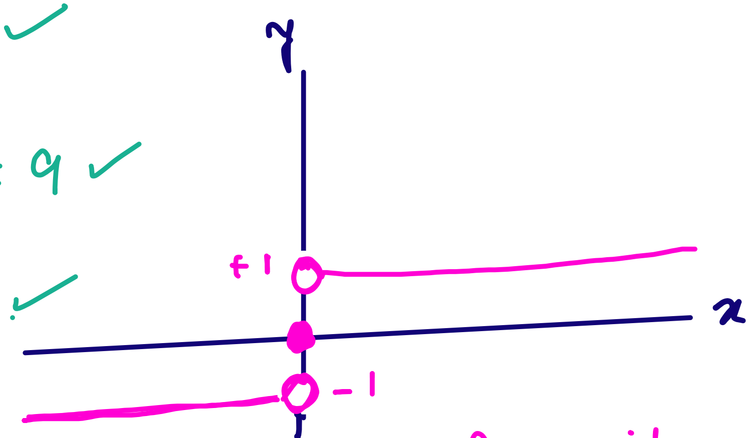
$\lim_{x \to 3^+} f(x) = 9$ ✓

$\lim_{x \to 3^-} f(x) = 9$ ✓

$f(3) = 9$ ✓



$f(x) = x^2$.



sign → $f(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0. \end{cases}$

Continuous vs. Discontinuous

Continuous: Can be drawn without lifting the pen from the paper.

A function is continuous at a point if and only if

$$\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x) = f(a).$$

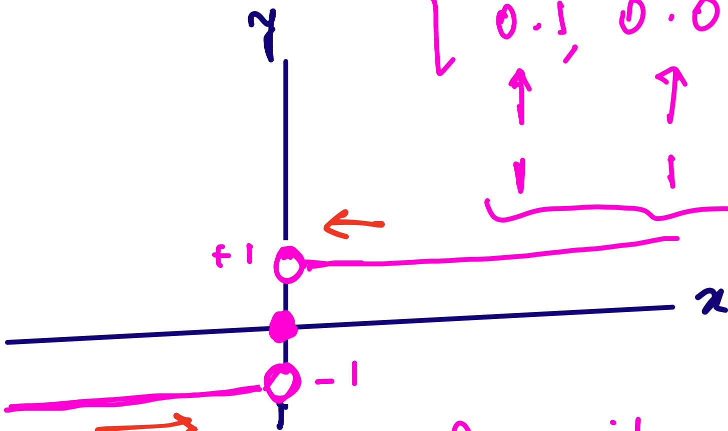
If any of these are not the same, the function is discontinuous.

Discontinuous: Cannot be drawn without lifting our pen from the paper.

Limits.

Right-hand limit. $\lim_{x \to 0^+} f(x) = 1.$

$[0.1, 0.01, 0.001, 0.0001, 0.00001 \ldots] \longrightarrow$



$$f(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0. \end{cases}$$

sign

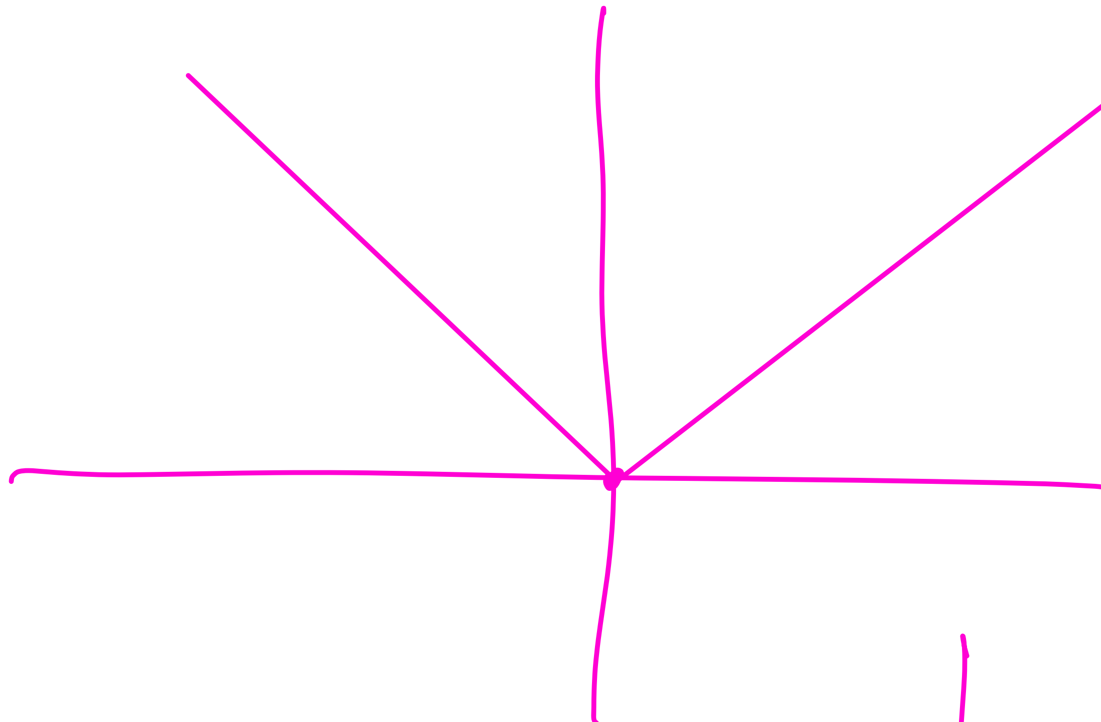
$[-0.1, -0.01, -0.001, \ldots \longrightarrow$

Left hand limit. $\lim_{x \to 0^-} f(x) = -1$

Practice: Are the following functions continuous or discontinuous?

Function	Domain	Range	Continuous?
① $f(x) = \frac{1}{x}$	$(-\infty, \infty)$	$(-\infty, \infty)$.	Discontinuous. at $x = 0$.
② $f(x) = e^x$	$(-\infty, \infty)$	$(0, \infty)$	continuous.
③ $f(x) = \lvert x \rvert$	$(-\infty, \infty)$	$(0, \infty)$	continuous.
④ $f(x) = \log(x)$	$(0, \infty)$	$(-\infty, \infty)$	continuous
⑤ $f(x) = \frac{1}{1+e^{-x}}$	$(-\infty, \infty)$	$(0, 1)$	continuous.
⑥ $f(x) = \cos(x)$	$(-\infty, \infty)$	$(-1, 1)$	continuous
⑦ $f(x) = \sin(x)$	$(-\infty, \infty)$	$(-1, 1)$	continuous
⑧ $f(x) = \tan(x)$	$(-\infty, \infty)$	$(-\infty, \infty)$	Discontinuous at $x =$ odd multiples of $\pi/2$.

$x = \frac{(2n-1)\pi}{2}$



Statistics

* Two samples, ~1M records.

store_id, product_id, quantity_sold

3 50 (num. purchases)

store_id, product_id, quntity_audited.

1 3 50.

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

where a^2, b