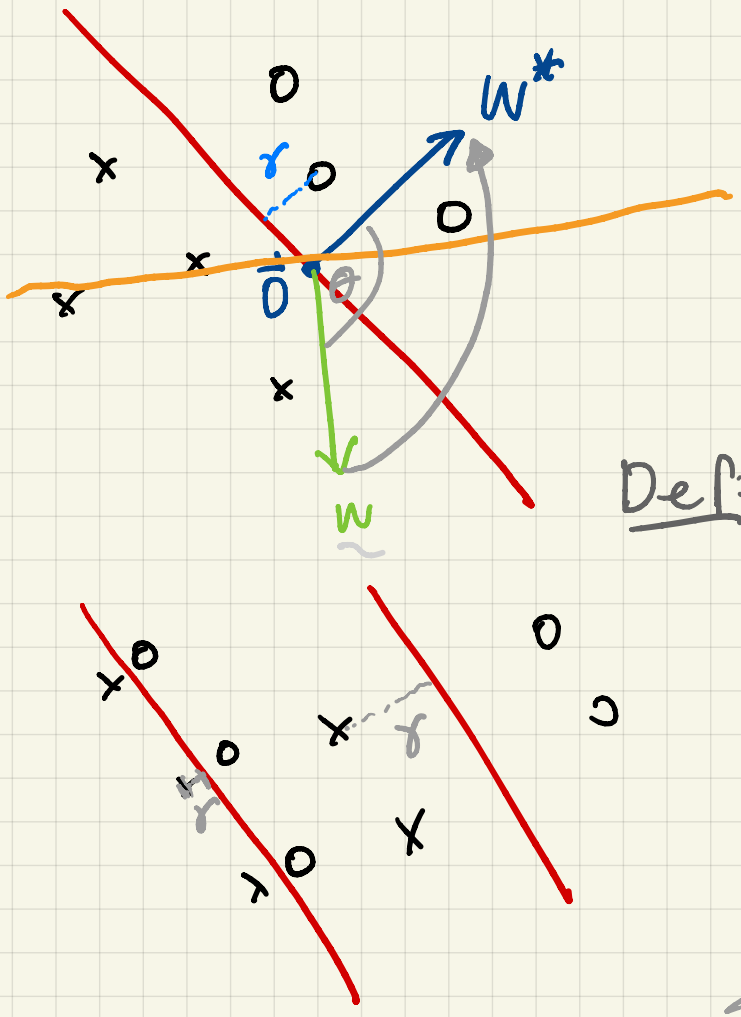
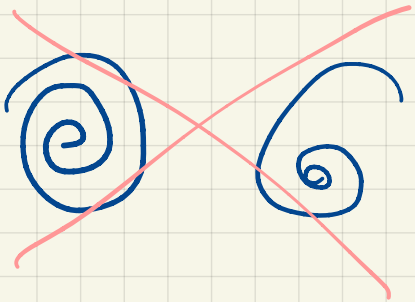


$w \rightarrow w^*$ in at most $\frac{1}{\gamma^2}$ updates

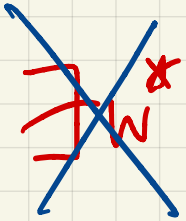
Def: γ is the smallest distance between point and the hyperplane defined by w^*, defined as

$$\gamma = \min_{(x,y) \in D} \left| x^T w^* \right|$$

Nerve cell → (Perceptrons) → Neural Networks → Deep Learning



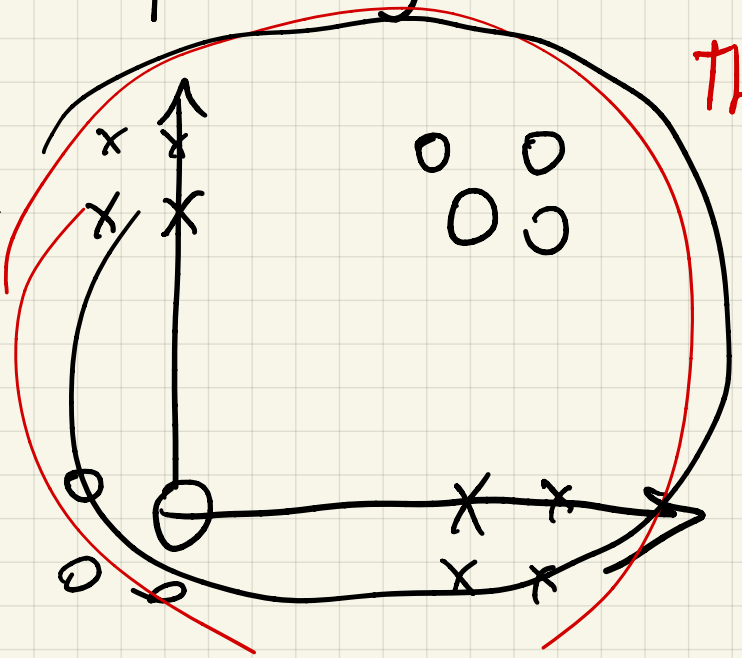
Limitation ?



The Perceptrons

by Minsky & Papert 1969

AI Winter ⇒ Machine Learning

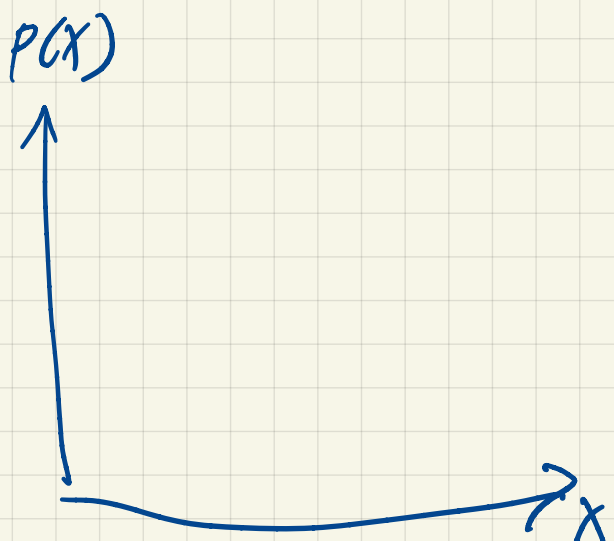


The XOR Problem

$$D = \{ (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \}$$

$(x_i, y_i) \sim P$ (inaccessible)

PDF / PMF



Probability distribution

Random experiment $\Longrightarrow$ outcomes (uncertain)

Ex. Coin toss $\Rightarrow$ $S = \{H, T\}$

$P(H) = 0.5$

$P(T) = 0.5$

$P(S) = 1$

- Probability Function → $P(X, Y)$
- Probabilistic Prediction
- Estimating Probability

Maximum Likelihood Estimation (MLE)

$$D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

Each $(x_i, y_i) \sim P$ (inaccessible)

$S = \{HH, TH, HT, HH\}$

$P : S \to R^+$

Probability distribution = function that maps samspace to positive real number

Event

Random experiment → Random outcomes (uncertains)

Coin toss → H or T

$X \approx$

$P(HH) = 1/4$ $\quad P(TH) = 1/4$

$P(HT) = 1/4$ $\quad P(TT) = 1/4$

$P(\{HH, TH\})$

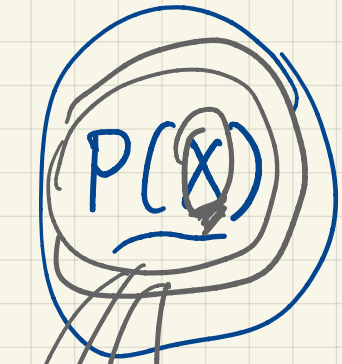
$= P(HH \text{ or } TH)$

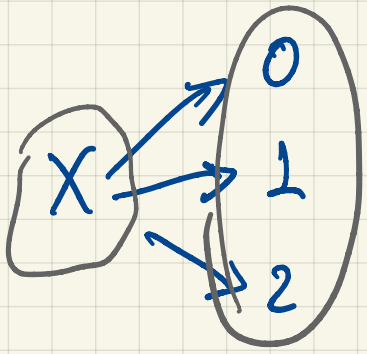
$= 1/4 + 1/4 = 2/4$

$X : S \to R$

X is a discrete R.V. denoting the number of heads in two throws of coin toss

unbias





PMF

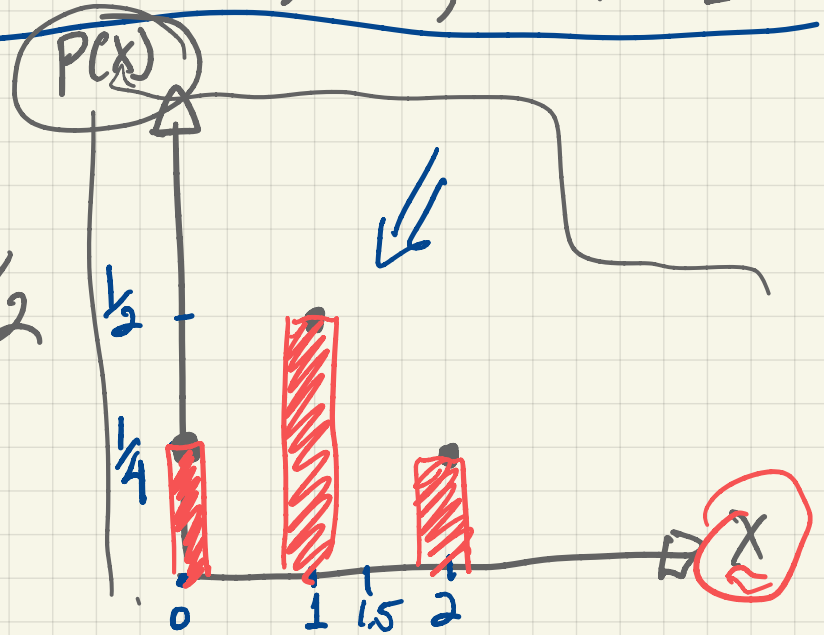
PDF

$P(X=0, X=1, X=2) = 1$

$P(X=0) = P(\{TT\}) = 1/4$

$P(X=1) = P(\{HT, TH\}) = 1/2$

$P(X=2) = P(\{HH\}) = 1/2$



Maximum Likelihood Estimation (MLE)

Background: Probability Theory

$$D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

Assumption: $(x_i, y_i) \sim P$

inaccessible distribution

$S = \{o_1, o_2, \ldots, o_n\}$

$P(o_i) > 0$

$P(o_i) \leq 1$

$P(o_1) + P(o_2) + \ldots P(o_n) = 1$

Probability: A function that maps each outcome to some positive real number

Probability distribution

Random experiment → outcomes (uncertain)

$P: S \to \mathbb{R}^+$

Two Coin tosses → $S = \{HH, HT, TH, TT\}$

sample space

$S = \{HH, HT, TH, TT\}$

$P(HH) = 0.25$

$P(HT) = 0.25$

$P(TH) = 0.25$

$P(TT) = 0.25$

Event $\subseteq$ Sample space

$P(\{HH, TT\}) = 0.5$

Event

n

~~Two~~ coin tosses $\longrightarrow$ $S = \{0, 1, 2\}$

X denotes the number of heads in two coin tosses

$P(X=0) = P(TT) = 0.25$

$P(X=1) = P(TH) + P(HT) = 0.5$

$P(X=2) = P(HH) = 0.25$

⋮

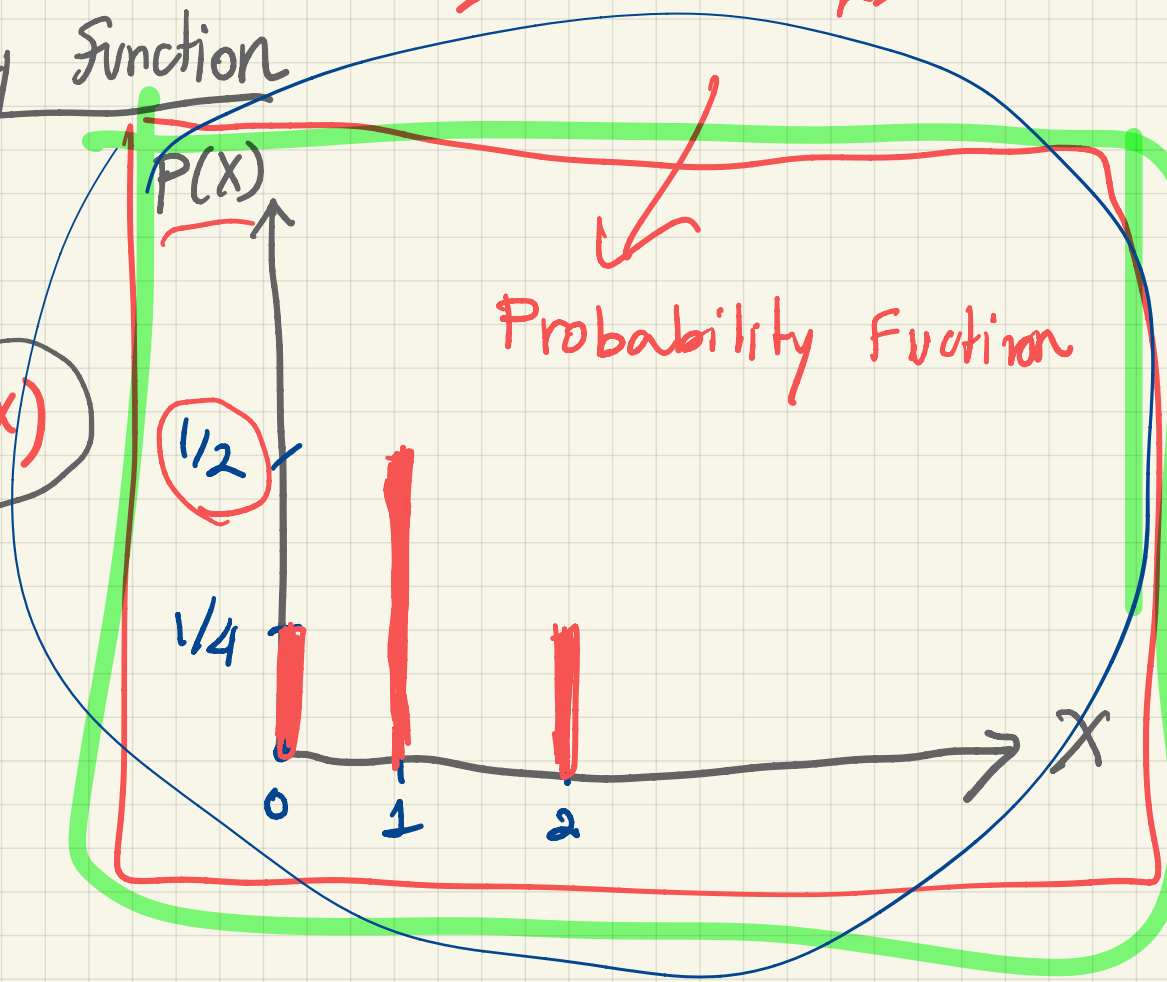
X

$S = \{0, 1, 2\}$

Discrete R.V.

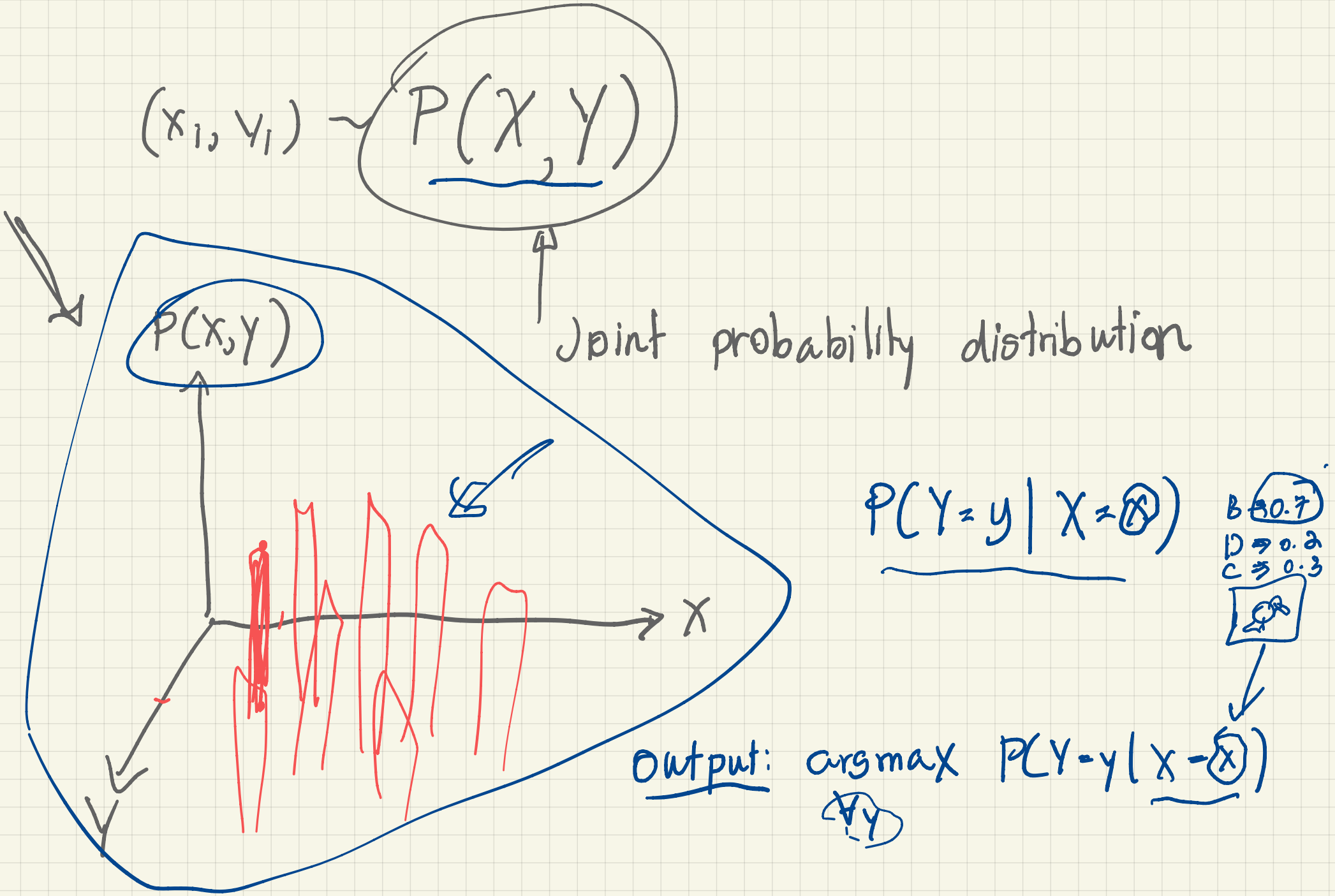
P(X) — Probability function

Probability distribution of P(X)



10 experiments → X = 1

5


$(x_1, y_1) \sim P(X, Y)$
Joint probability distribution
$P(X, Y)$
X
Y
$P(Y = y \mid X = x)$
B ⇒ 0.7
D ⇒ 0.2
C ⇒ 0.3
Output: $\arg\max_y P(Y = y \mid X = x)$