Interview Prep Sessions

[Linear algebra & optimisation]

Q: Which of the following lines are unique?

→ Automated | Easy

a) $y = 3x + 6$

b) $\frac{y}{18} - \frac{x}{6} = \frac{1}{3}$

c) $\frac{y}{3} - x + 2 = 0$

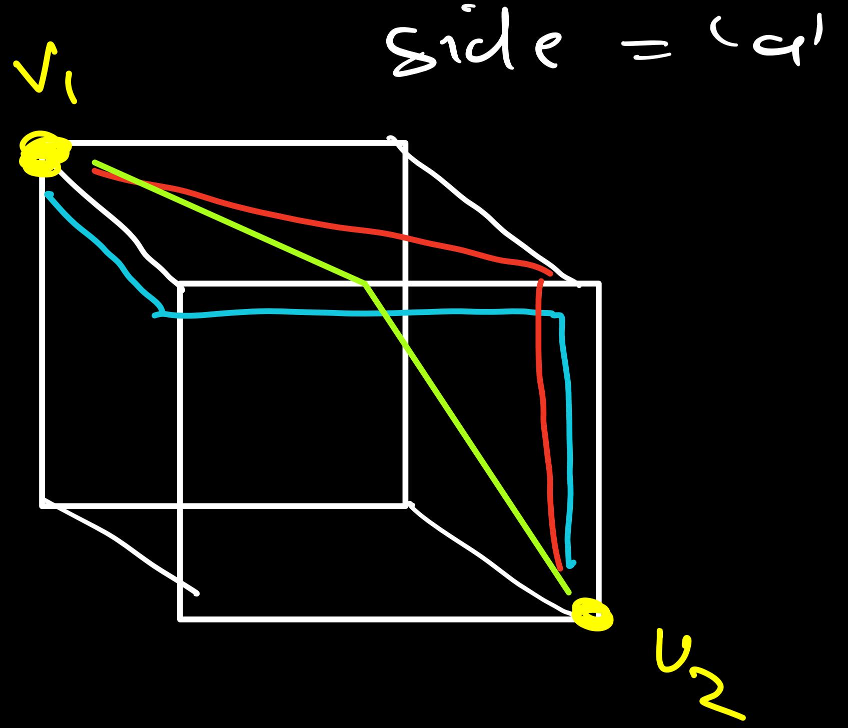
d) $0.33y - x - 2 = 0$

a, b, d are just rearrangements of each other.

c is unique

Q: Imagine an ant on a cube of side 'a'. What is the shortest distance for ant to travel from one vertex to the opposite vertex?

→ Live | Easy



→ a) $3a$ units

→ b) $\sqrt{2}a + a$

→ c) $2\sqrt{a^2 + \frac{a^2}{4}}$

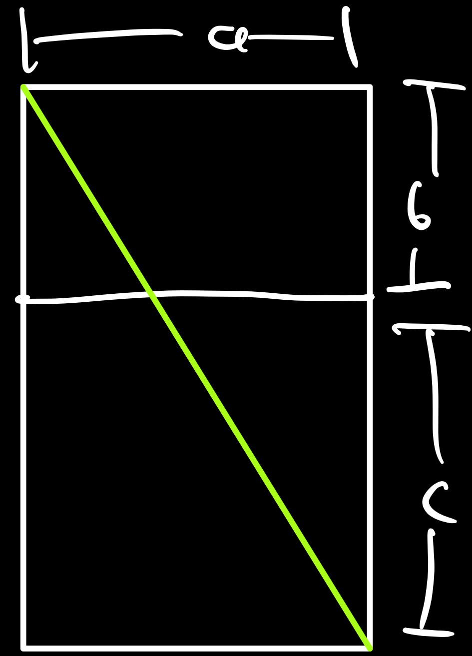
$= 2\sqrt{\frac{4a^2 + a^2}{4}} = \boxed{\sqrt{5}\, a}$

Variation / Follow up:

What is its a cuboid, dimensions

$a \times b \times c$?

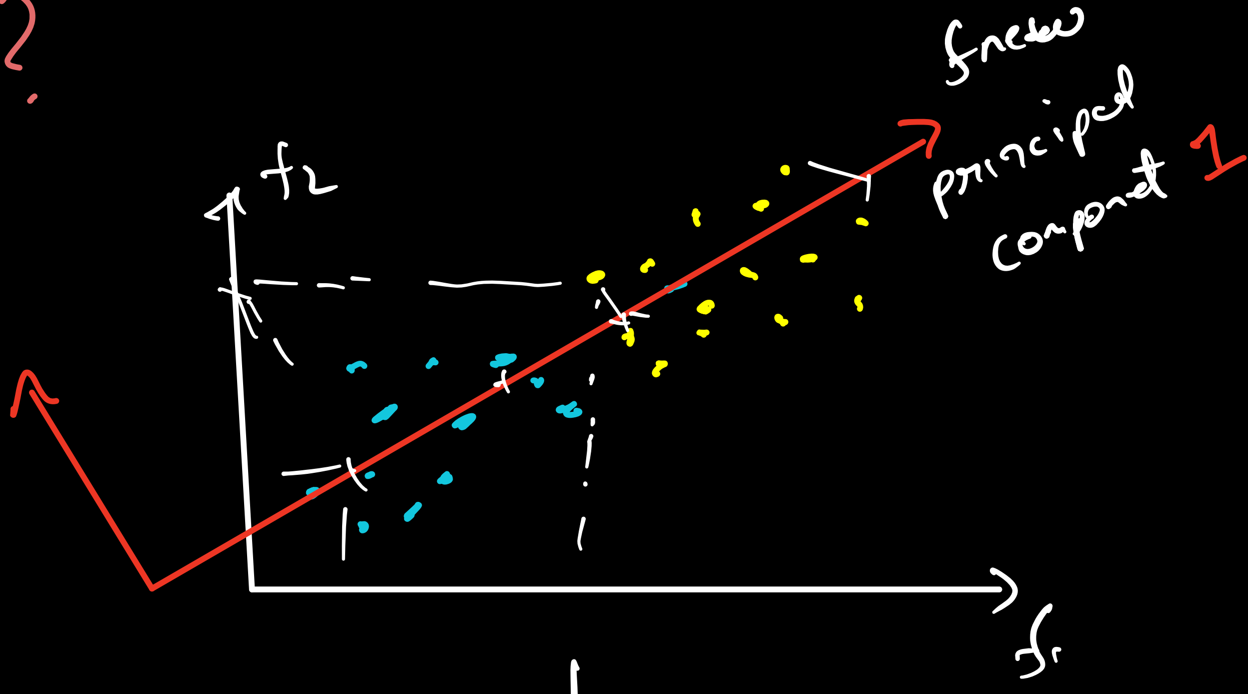
$\therefore \sqrt{a^2 + (b+c)^2}$



Q: Let's say I have n-dimensional data, and I projected [PCA] on 2D, in this new 2-D space I can easily seperate 2 classes using a linear decision boundary. Can you calculate the feature importance of the original features? if yes, how?

Live | Medium

the projected componed from PCA with largest

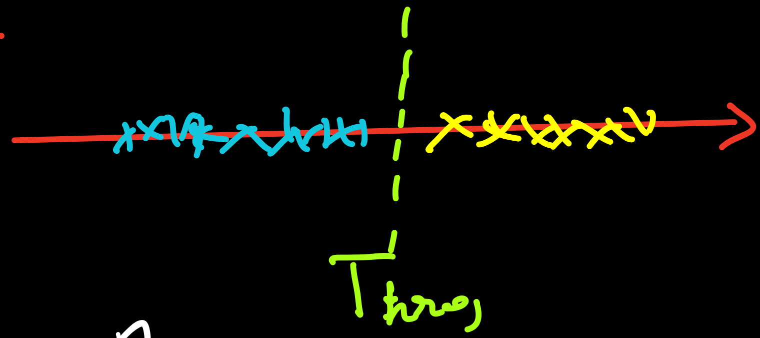


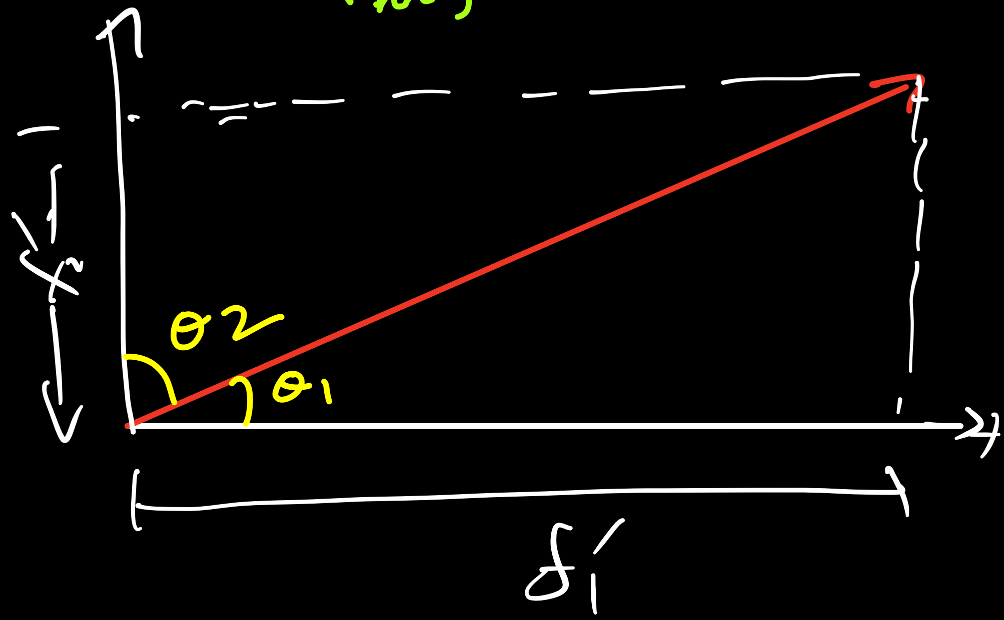
eigen value has 100%.

feature importance

Importance of orignal featurs.

- projet original axes to new axis
- Calculate the variance of respective projections
- FI can be





$\rightarrow \dfrac{f_1'}{f_1'+f_2'}, \dfrac{f_2'}{f_1'+f_2'}$ **OR** $\dfrac{\cos(\theta_1)}{\cos\theta_1+\cos\theta_2}, \dfrac{\cos\theta_2}{\cos\theta_1+\cos\theta_2}$

Q: Can you implement KNN, where all features are categorical, with each column having large number of categories?

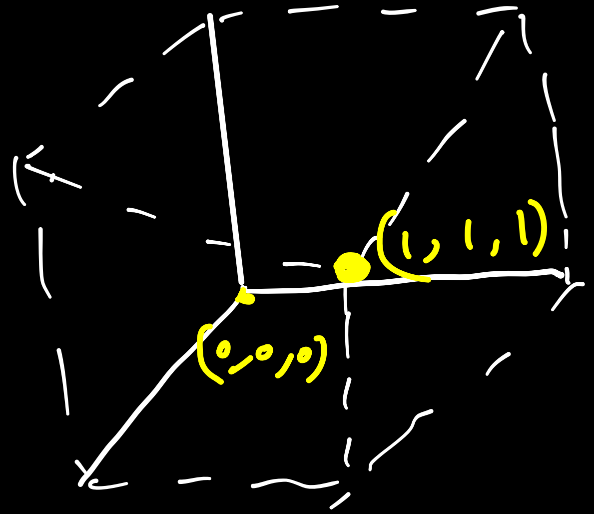
Note: Originally this question was asked for an unsupervised ranking problem so target encoding is not allowed.

Live | Hard

→ One-hot encoding would lead to too many dimensions

→ computing distances become hard.

Let's take example of 3 dimensions



Farthest point is

1, 1, 1

$$= \sqrt{1^2+1^2+1^2} = \sqrt{3}$$

expensive

2nd farthest points

0,1,1, 1,1,0, 1,0,1

$$= \sqrt{0^2+1^2+1^2} = \sqrt{2}$$

shortest

= 0,0,1, 1,0,0, 0,1,0

= 1 unit

What if I just take the sum of coordinates

$\Sigma 0, 0, 1 = 1 \to$ shortest

$\Sigma 0, 1, 1 = 2 \to 2^{nd}$ far

$\Sigma 1, 1, 1 = 3 \to$ farthest

order of distance is not affected.

For ranking problem

distance from $\vec{a}$ to $\vec{b}$

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 \cdots ( )}$$

$\to$ euclidean

$\times$ slow

a = 0 1 0 1 1 0 0

b = 1 1 0 1 0 1 1

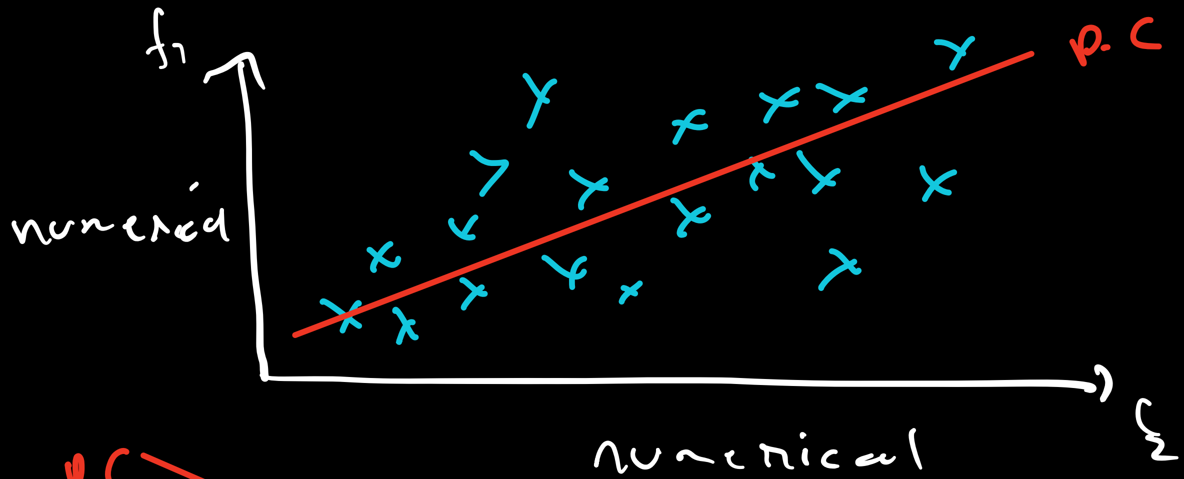
$\to$ XOR operator + sum

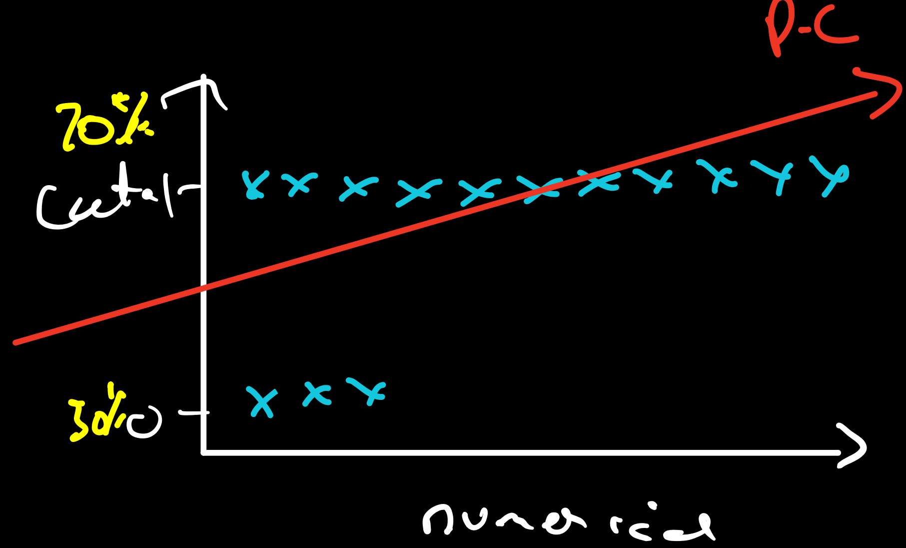
sum(abs(a − b))

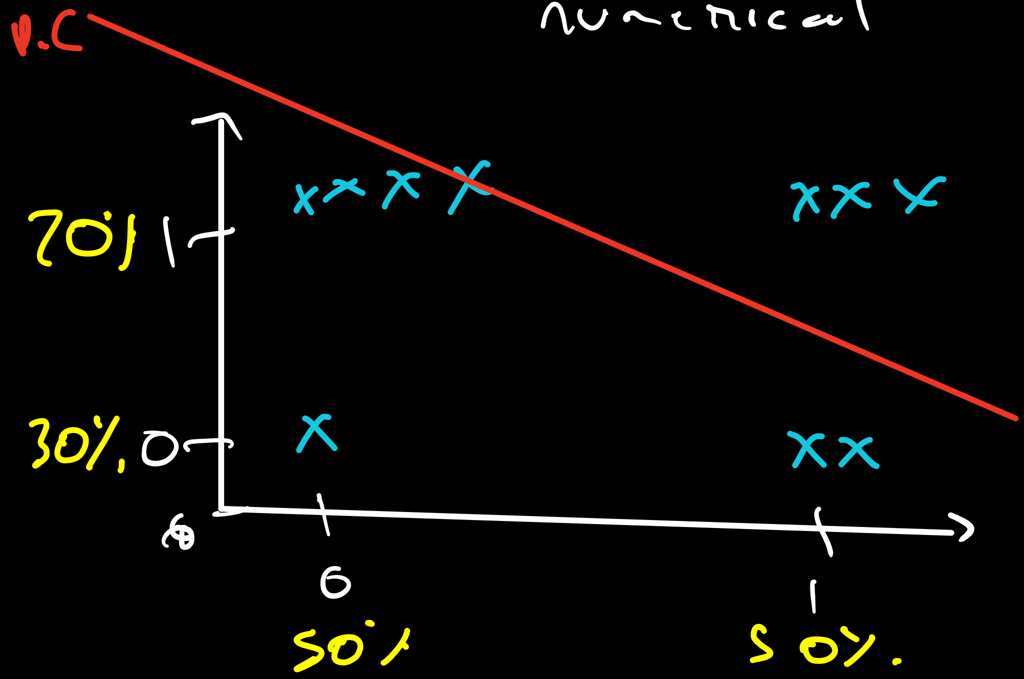
$\to$ Hamming distance

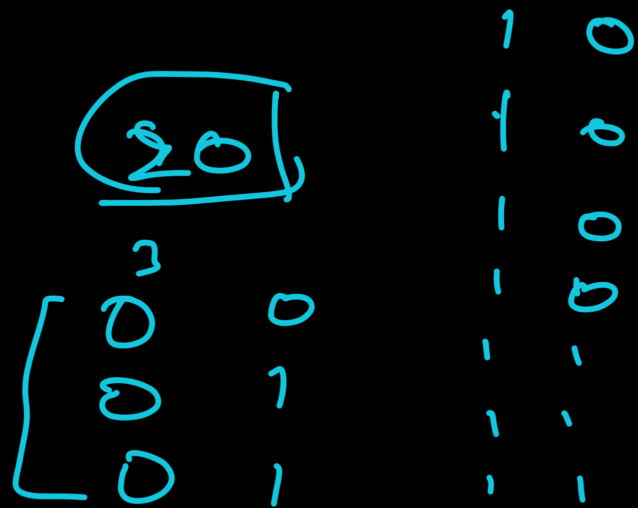
Q: Can you use PCA on categorical variables?

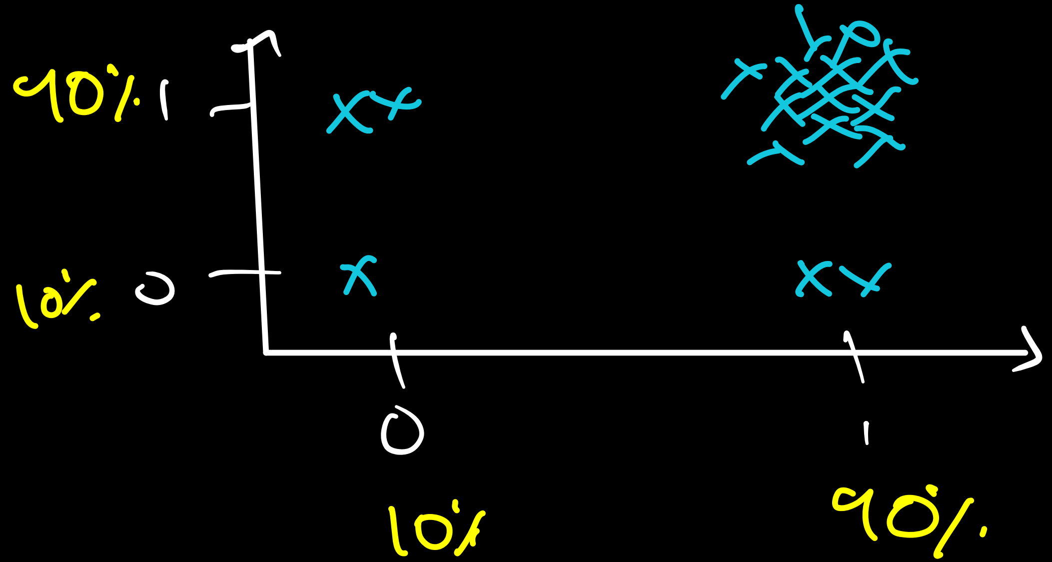
→ Live | Discussion











P.C ??

→ some 3rd variable

It may be useful for a large number of categorical variables or mix with numerical variables, since variance of some boolean variable might be very low.

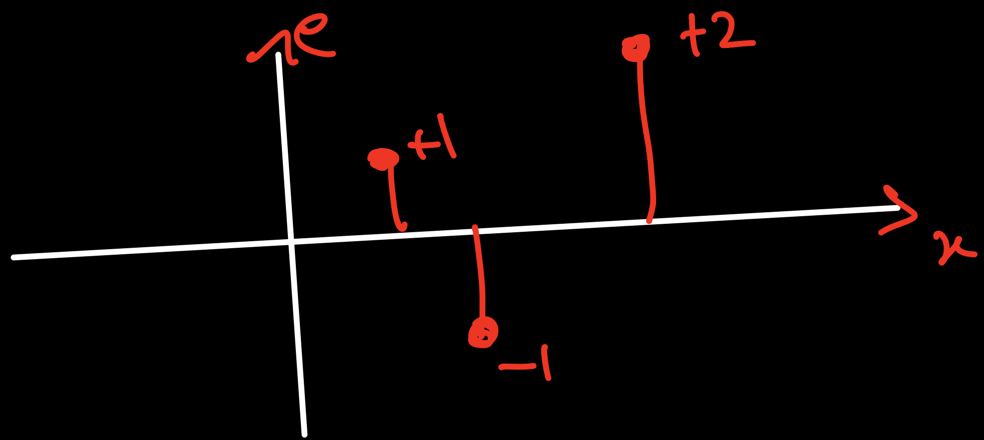
Q: What are the reasons to choose SSE as loss function for regression?

→ Differentiable

→ Convex

→ penalizes 1 large error more than 2 small errors

Eg:



$|1| + |1| = 2$

$|2| = 2$ same

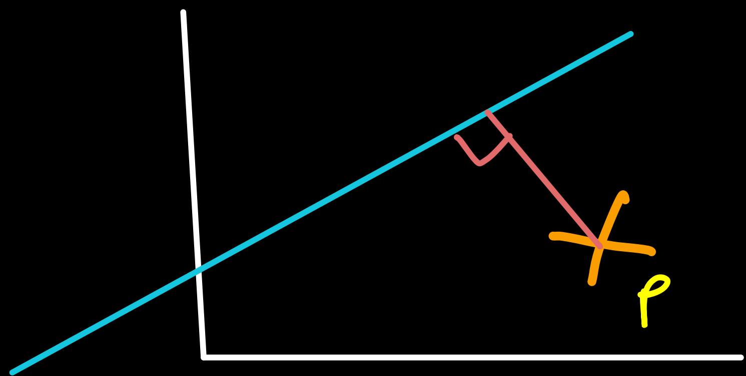
Square instead of | |

$1^2 + 1^2 = 2$

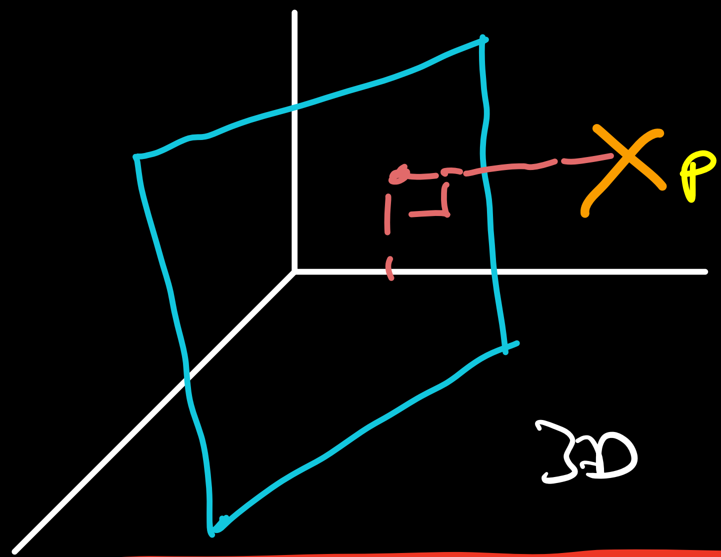
Higher penalty → $2^2 = 4$

Q: Write a program to find the distance of a point from a hyperplane.

Live | Hard



2-D



3D

Let p be represented by $\rightarrow$ $\vec{p} = [x_1, x_2, x_3 \cdots x_n] = \vec{x}$

Skip

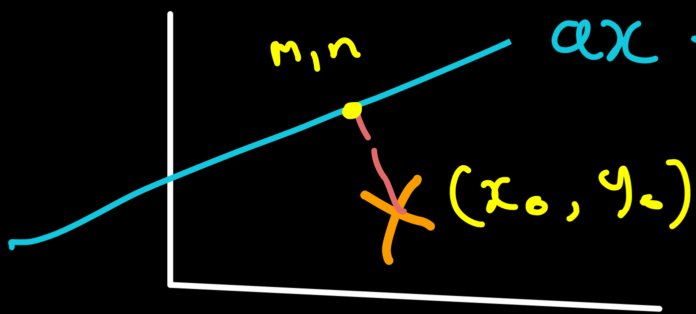
Hard

and hyperplane as:

$$w_1 x_1 + w_2 x_2 + w_3 x_3 + \dots w_n x_n + w_0 = 0$$

$$\rightarrow \quad w^T x + w_0 = 0$$

Let's take case of 2D



$$ax + by + c = 0$$

$$\downarrow$$

$$\text{slope} = y = \underbrace{\boxed{-\frac{a}{b}}}_{m} x - \frac{c}{b}$$

∴ perpendicular slope

$$= m_1 \cdot m_2 = -1$$

$$\therefore \; -\frac{a}{b} \times m_2 = -1$$

$$\therefore \; m_2 = \frac{b}{a}$$

$\therefore$ perpendicular line

$$\rightarrow \quad y = \frac{b}{a} x - \underbrace{K}_{??}$$

$\perp$ line needs to pass through the point

$$\rightarrow \quad y_1 = \frac{b}{a} x_1 - K$$

$$\rightarrow \quad K = \frac{b}{a} x_1 - y_1$$

$$\therefore \perp \text{ line} = y = \frac{b}{a} x - \left(\frac{b}{a} x_1 - y_1\right)$$

At point of intersection both coordinates will be equal.

$$l_1: \quad y = -\frac{a}{\ } x - \frac{c}{\ }$$

L_2: $y = \frac{b}{a}x - \left(\frac{b}{a}x_1 - y_1\right)$

$\therefore -\frac{a}{b}x - \frac{c}{b} = \frac{b}{a}x - \frac{b}{a}x_1 + y_1$

$\therefore -x\left(\frac{a}{b} + \frac{b}{a}\right) = y_1 + \frac{c}{b} - \frac{b}{a}x_1$

$\therefore -x\left(\frac{a^2+b^2}{\cancel{ab}}\right) = \frac{y_1 ab + ac - b^2 x_1}{\cancel{ab}}$

$\therefore \quad x = \frac{b^2x_1 - aby_1 - ac}{a^2+b^2}$

$$y = mx + c$$

$$y = \frac{-1}{m}x + K$$

$$y_1 = \frac{-1}{m}x_1 + K$$

$$K = y_1 + \frac{x_1}{m}$$

$$\therefore \; y = \frac{-x}{m} + \frac{y_1 m + x_1}{m}$$

$$\therefore \quad mx + c = \frac{-x}{m} + \frac{y_1 m + x_1}{m}$$

$$x\left(m + \frac{1}{m}\right) = \frac{y_1 m + x_1 + cm}{m}$$

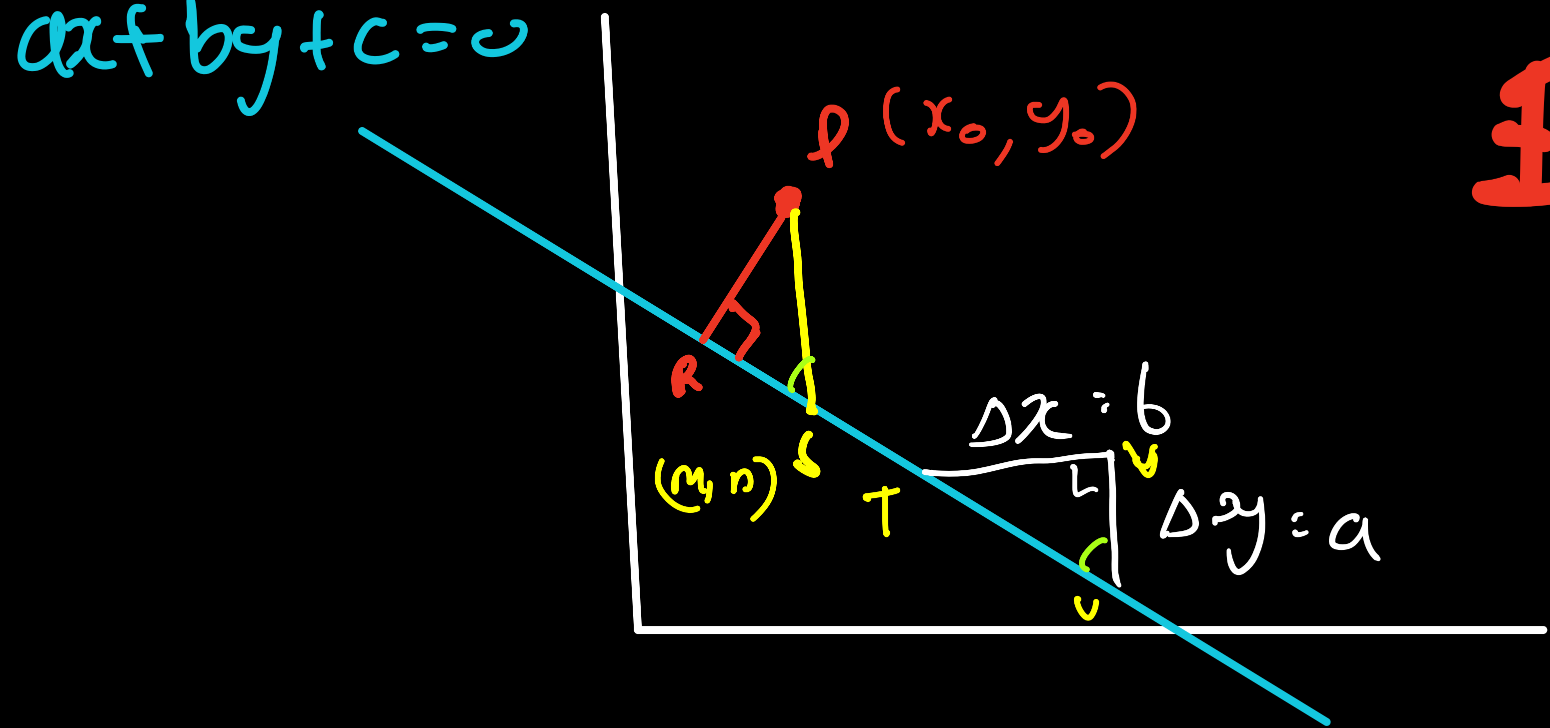
$$x(m^2 + 1) = y_1 m + cm + x_1$$

$$\therefore \quad x = \frac{y_1 m + cm + x_1}{(m^2 + 1}$$

$$\therefore \quad y = m\left(x\right) + c$$

$$d_1 - \quad \frac{a x_0 + b y_0 + c}{\sqrt{a^2 + b^2}} \quad ?$$

EASIER



$$\therefore \frac{\overline{PR}}{\overline{PS}} = \frac{\overline{TV}}{\overline{TU}}$$

$$\therefore \overline{PR} = \frac{|y_0 - n|\,|b|}{\sqrt{a^2+b^2}}$$

$$\therefore \frac{\overline{PR}}{y_0 - m} = \frac{|b|}{\sqrt{a^2+b^2}}$$

$\therefore$ Also, $am + bn + c = 0$

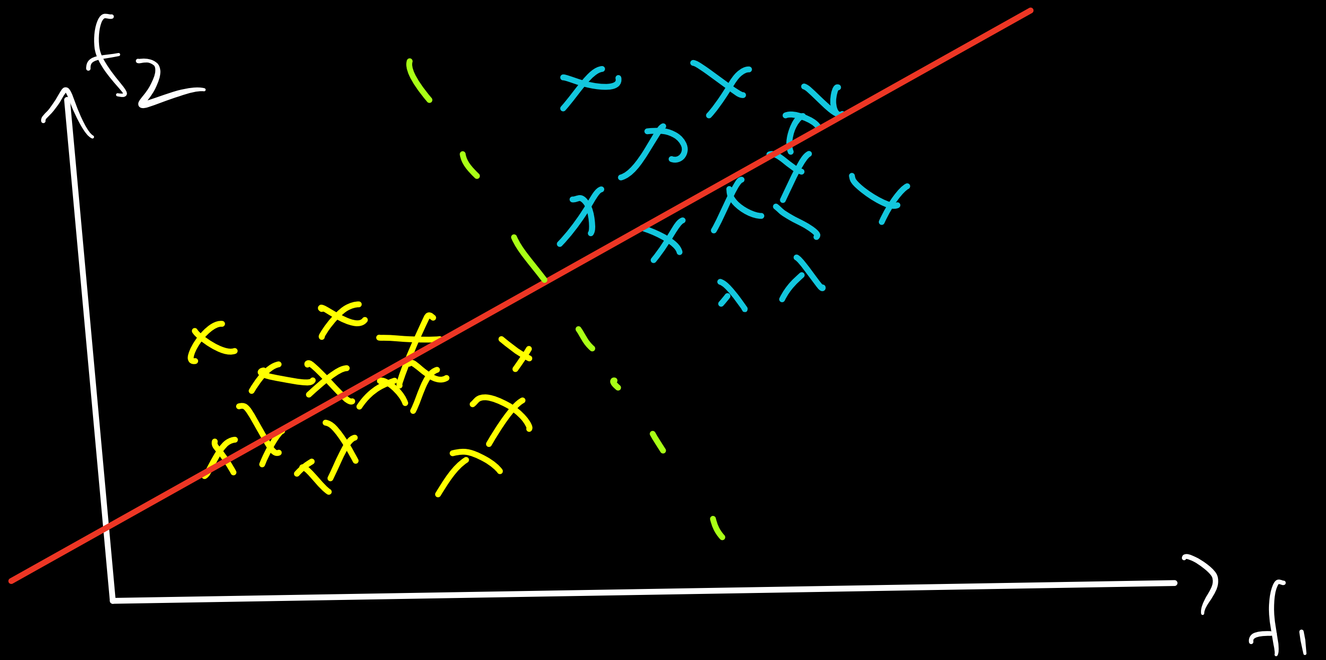
$$\therefore n = \frac{-am - c}{b} \qquad \# \; m = x_0$$

$$\therefore \overline{PR} = \frac{by_0 + am + c}{\sqrt{a^2 + b^2}}$$

$$= \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}$$

$\therefore$ in general $\rightarrow$ $$\frac{w^T x_0 + c}{|w|}$$

Q: Can you change the cost func^n of PCA and make it a supervised learning technique? [classification]



Simple PCA, i.e var max may give good results directly, but

You may want a line $\perp$ to decision boundary.

$\therefore$ PC $\longrightarrow$ $ax + by + c = 0 \perp$ DB

obtained via Log Reg

OR

max projected $\dfrac{\text{inter class var}}{\text{intra class var}}$

$\rightarrow$ LDA (Linear discriminant analysis)

So for all possible slopes and intercepts we want the line such that, after projecting points on the new line, we get max inter class variance and min intra class variance.

→ If you know the math for projections, you can frame this as an opt problem and solve using matrix factorisation OR gradient descent

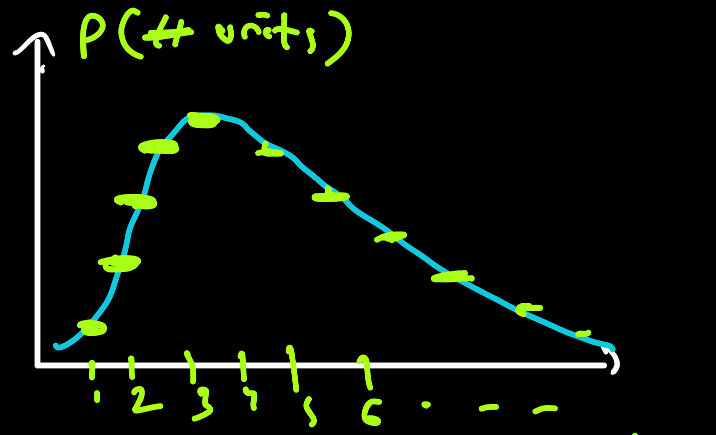
Q: News vendor problem

A new vendor needs to decide how many newspapers to buy in the morning. If he sells them, he makes a profit

cp = cost price

sp = selling price

s = salvage price



$Q+1$ units will be sold

if demand $\geq Q+1$

1 unit is wasted when demand = Q units

profit = $sp - cp$ (when sold)

loss = $cp - s$ (when unsold)

if I produce too mony, i loose on potential profits

Opportunity cost

→ # extra units that could sell × (sp - cp)

× probability of selling them

If I am not able to sell some units I will loose my investment

Salvage loss →

unsold goods × (cost price - s)

× probability of not selling

Let's say we are adding 1 more unit to our current number of units Q

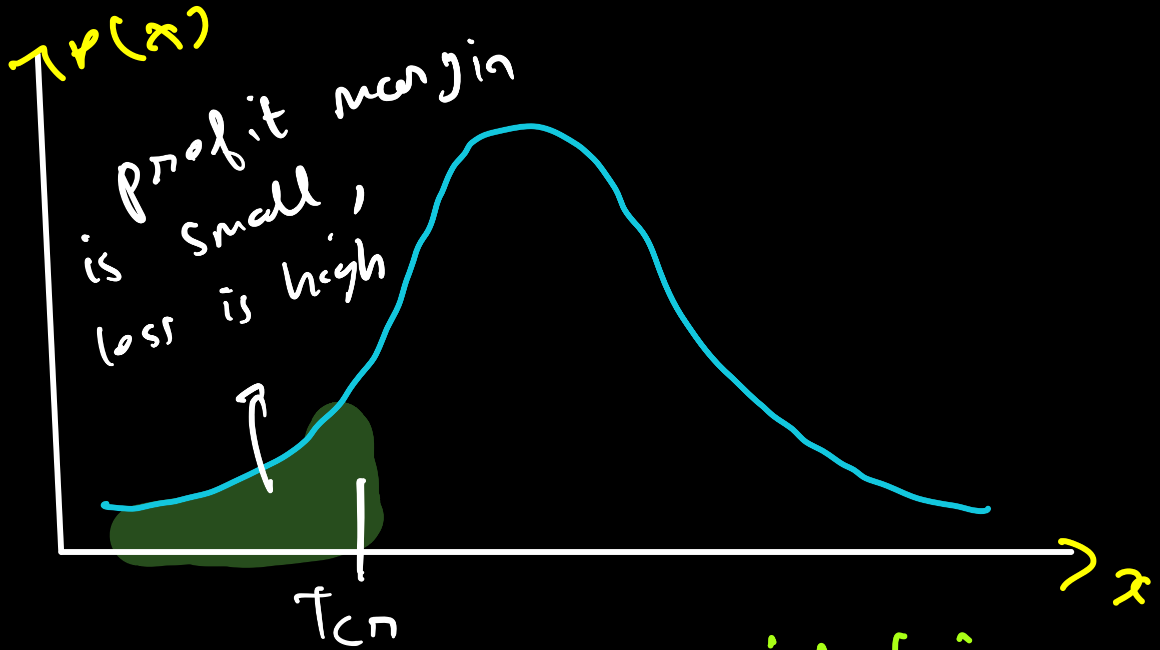
$$\underbrace{p(x \geq Q+1)}^{*} \, \underline{\underline{(sp-cp)}} = p(x \leq Q)(cp-s)$$

$$[1-cdf(Q)][sp-cp] = cdf(Q)(cp-s)$$

$$\frac{1-cdf(Q)}{cdf(Q)} = \frac{cp-s}{sp-cp}$$

$$\frac{1}{cdf(Q)} - 1 = \frac{cp-s}{sp-cp}$$

$$\therefore \frac{1}{cdf(Q)} = \frac{cp-s}{sp-cp} + 1 = \frac{sp-s}{sp-cp}$$



demand distribution

demand distribution

$$\therefore \quad cdf(Q) = \frac{sp - cp}{sp - s} = \underbrace{\frac{\text{profit}}{\text{profit} + \text{salvage}}}_{\text{critical ratio}}$$

$$\therefore \quad Q = cdf^{-1}\left(\frac{p}{p+s}\right)$$

Q: You are the lead data scientist for a retail company, How would you decide how much discount to offer to the customers?

Goal: → Maximise profit

$\neq$ revenue ← Note

$$\text{profit} = \#\text{units}\,(sp - cp)$$

$$\text{profit margin} = \frac{sp - cp}{cp}$$

$$\text{discount} = d\ \%$$

$$\text{discounted pm} = \frac{(1-d)(sp) - cp}{cp}$$

↳ can this be −ve ? → Yes

$d \in$?? $[0, 1]$

So profit margin depends on discount

What else will happen?

→ higher number of units will be sold

Let's say the % increase in sales is given by $\underline{s}$

How will you determine s ?

→ Get historic non-discounted sales per day.

→ Get " discounted sales per

day with corresponding discount

→ Learn a polynomial regression of discount vs % increase in sale

→ or some other complex ML model

→ other features:

→ discount type

→ product category (other attributes)

$\therefore S = f(d)$

↳ learn using ML / analysis

New total profit

% discount

avg selling price

$$P = n(1+S)\left[\frac{(1-d)(SP) - CP}{CP}\right]$$

units sold w/o discount

% of increase in sales due to discount

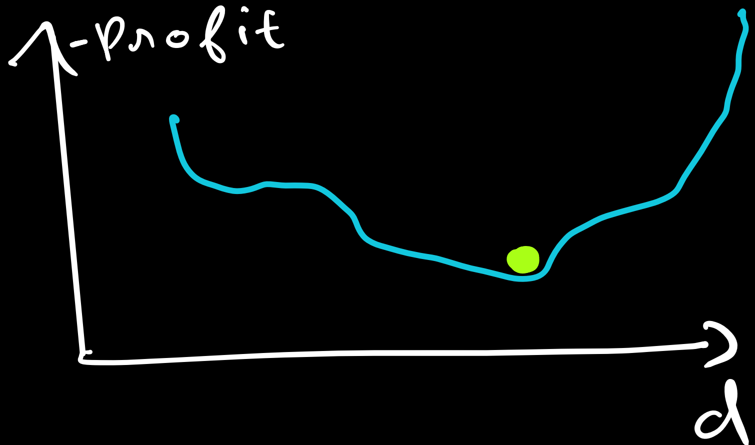
avg cp

$$\max_{d} : \phi(d) = n(1 + \underbrace{f(d)}_{\text{take a differentiable ML model}}) \left[\frac{(1-d)(sp) - cp}{cp} \right]$$

find $\frac{\Delta P}{\Delta d}$

↳ Throw in a gradient descent.

$$= \min_{d} : -\phi(d)$$



Is this all ?? → Industry knowledge / imagination / maturity

→ salvage price

→ shelf cost

→ Fix cost [electricity / staff / rent]

→ inventory turns

∴ if salvage price = sal (CP)

break even discount = $(1-d)\,SP = CP$

$$\therefore d_{bn} = \frac{SP - CP}{SP}$$

$$\Rightarrow (1 - d)\,SP = (sal)(CP)$$

$$d_{sal} \Rightarrow (1 - d_{sal})\,sp = (sal)(cp)$$

$$\therefore \quad d_{sal} = \frac{sp - sal(cp)}{sp}$$

$$shelf \propto \frac{1}{\#\,sales} = f_2(s^{-1})$$

$\hookrightarrow$ ML model 2 / analysis

$$\text{and} \quad s = f(d) \rightarrow \therefore shelf = f_2(f(d)^{-1})$$

$$\phi(d) = n(1 + f(d))\left[\frac{(1-d)(sp - cp)}{cp}\right] - \text{fixed} - f_2(f(d)^{-1})$$

$$d \in [0, d_{sal}]$$

Much more complex for inventory turns

Q: Can you convert an airline ticket pricing problem to an optimisation problem?

→ fixed costs (emis, seats, airport fee, staff, etc)

→ fuel $\propto$ (number of passengers)

⇒ profit $= n(1 + f(\text{price})).\text{price} - \text{fuel}(f(\text{price})) - \text{fixed}$

→ effect of competitors

→ forecast of demand [v. imp]