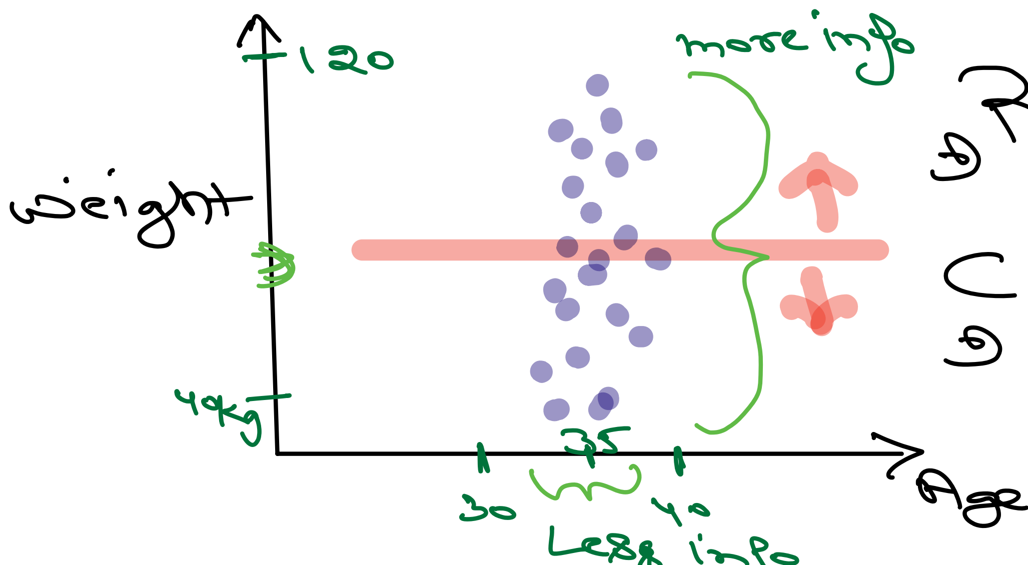
Why do we need to Reduce Dimension

⇒ By reducing dimensions to 2D or 3d, we will be able to visualize:

Clustering ⇒ That can help in finding optimal number

Classification & Regression ⇒ Reduce Training and inference Time Resources

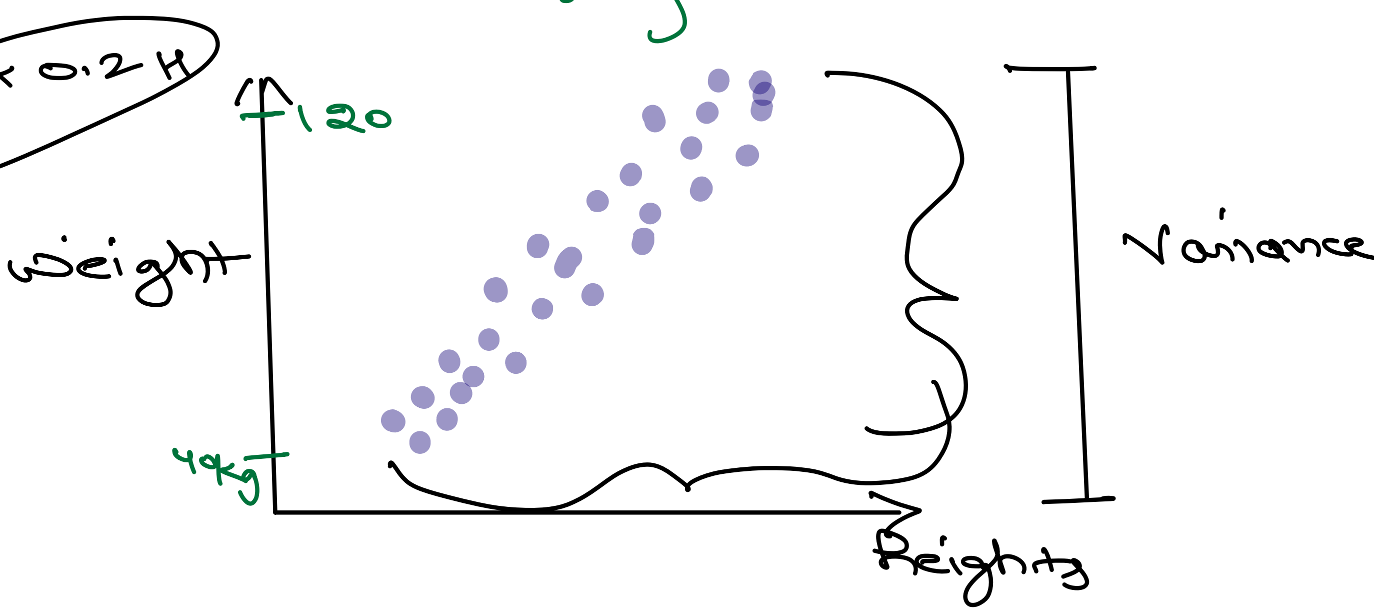
Principal Component Analysis



⇒ Regression Sugar level

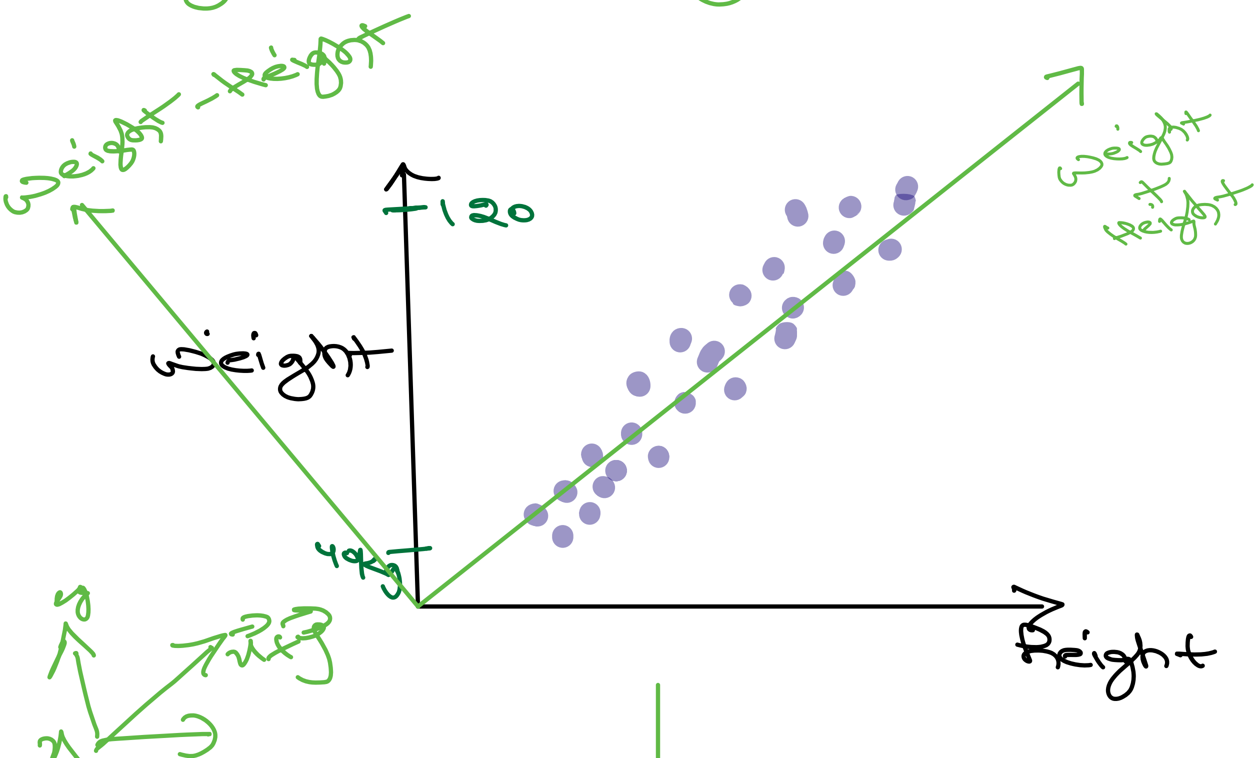
Classification ⇒ P Diabetic

$0.8\,W + 0.2\,H$

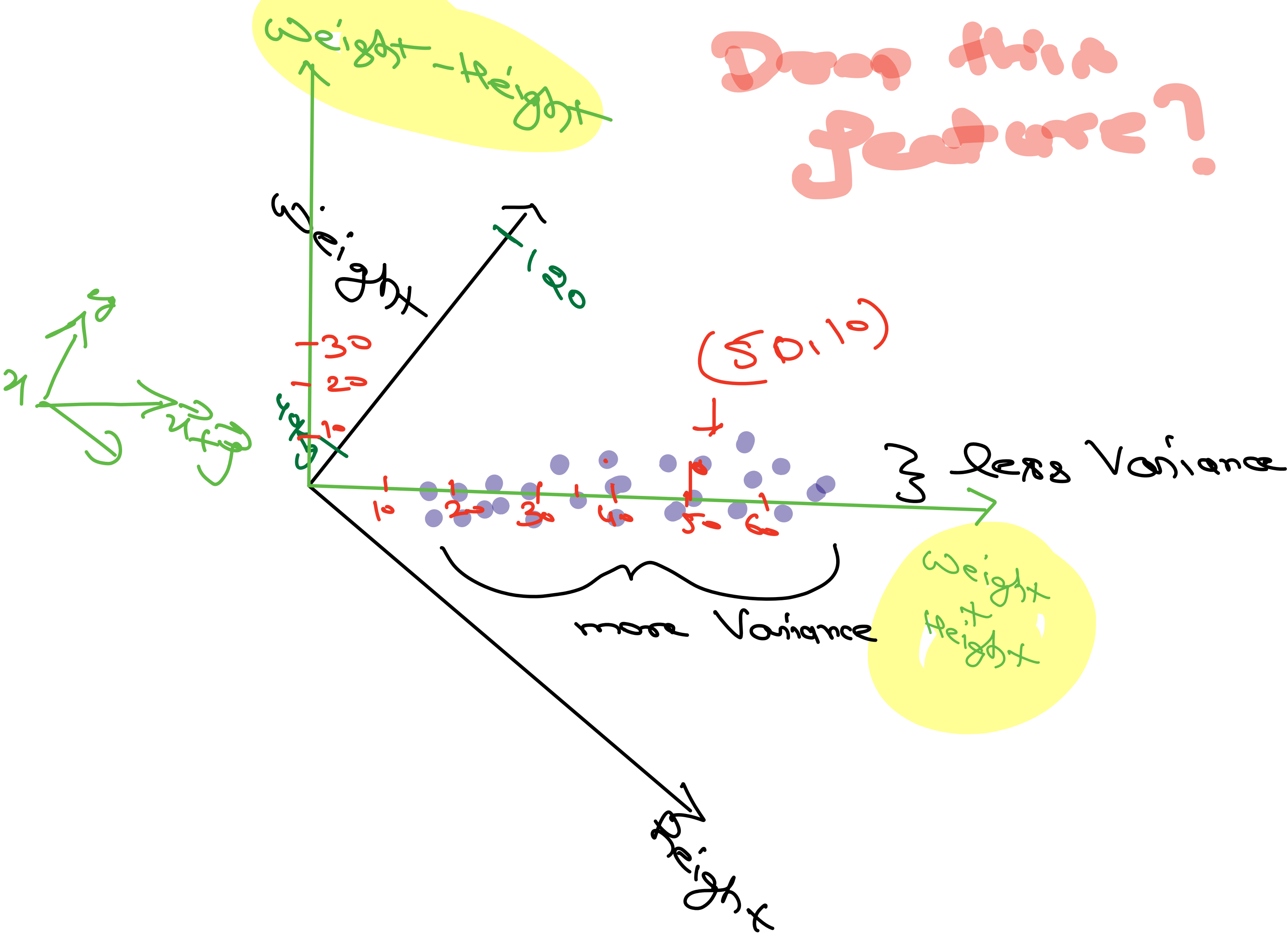


Which feature has more info?

1) Height ✓
2) Weight ✓
3) Height and Weight ✓



Linear axis Transformation



Conclusion:

1) Not all features are important
2) We can create new feature with Linear Combination of Existing feature
3) We can drop not so important features at the risk of losing some information.

m-features $\xrightarrow{\text{PCA}}$ m-features

100% Variance

Not all features in these m will be equally important

$m' < m$

m' can cover maybe 90-95% Variance from Original Dataset

w and h $\Rightarrow$ 2

Linear Combination

$w-h$ and $w+h$ $\Rightarrow$ 2

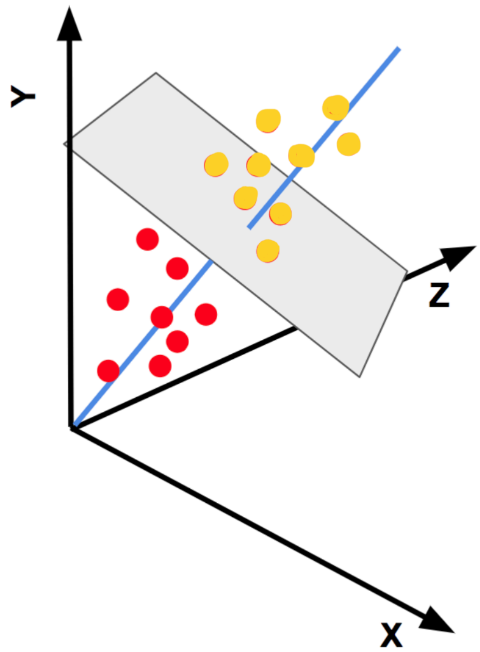
(5%) (95%)

Drop this and $m' \Rightarrow 1$

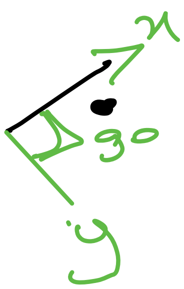
Variance $\Rightarrow$ 95%

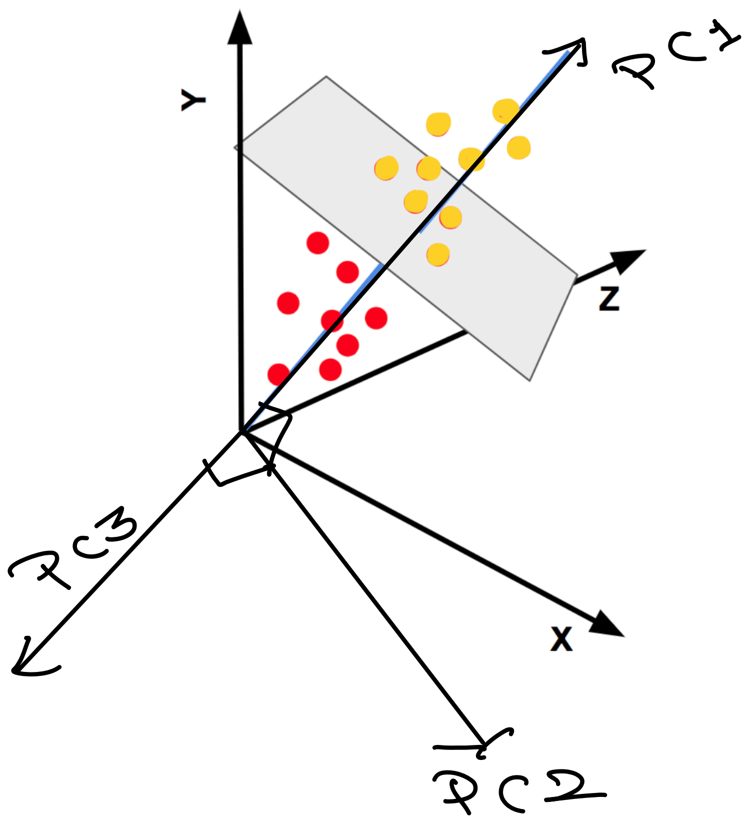
Goal of D.R

We want to Reduce Dimension while preserving as much Variance as possible



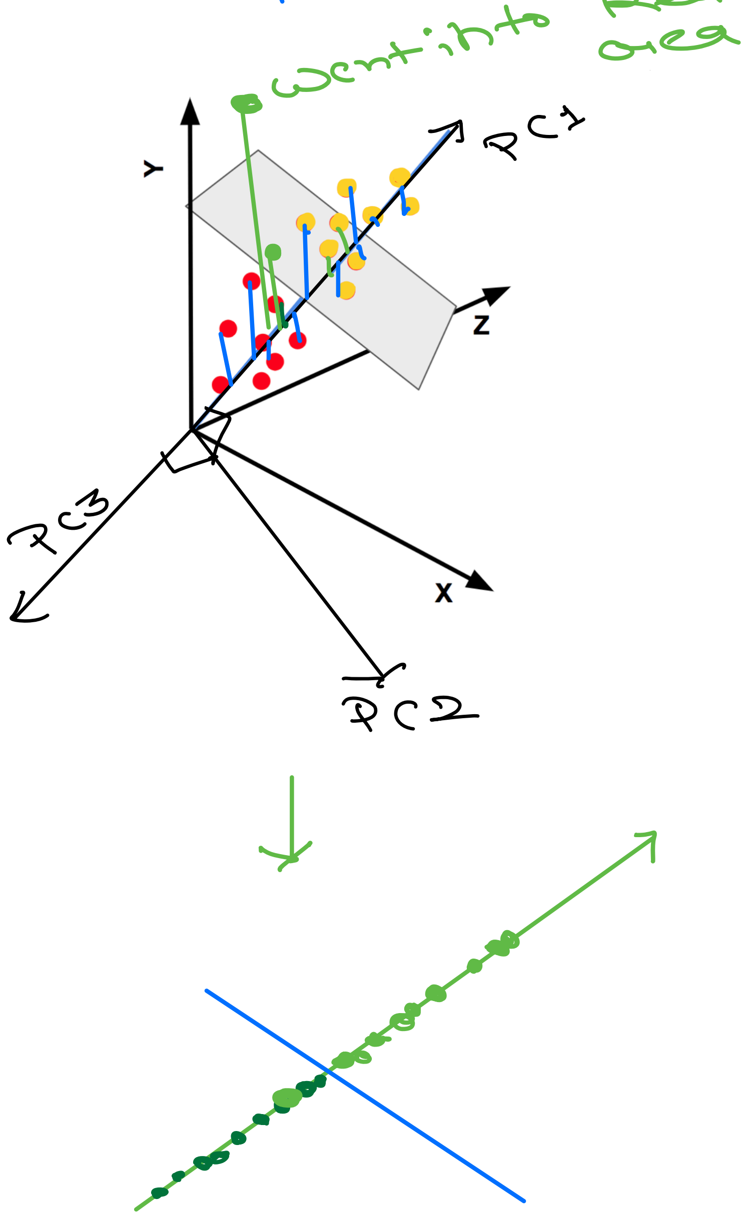
if we apply PCA or (x, y, z)
we get
$\Rightarrow$ 3 principal Component



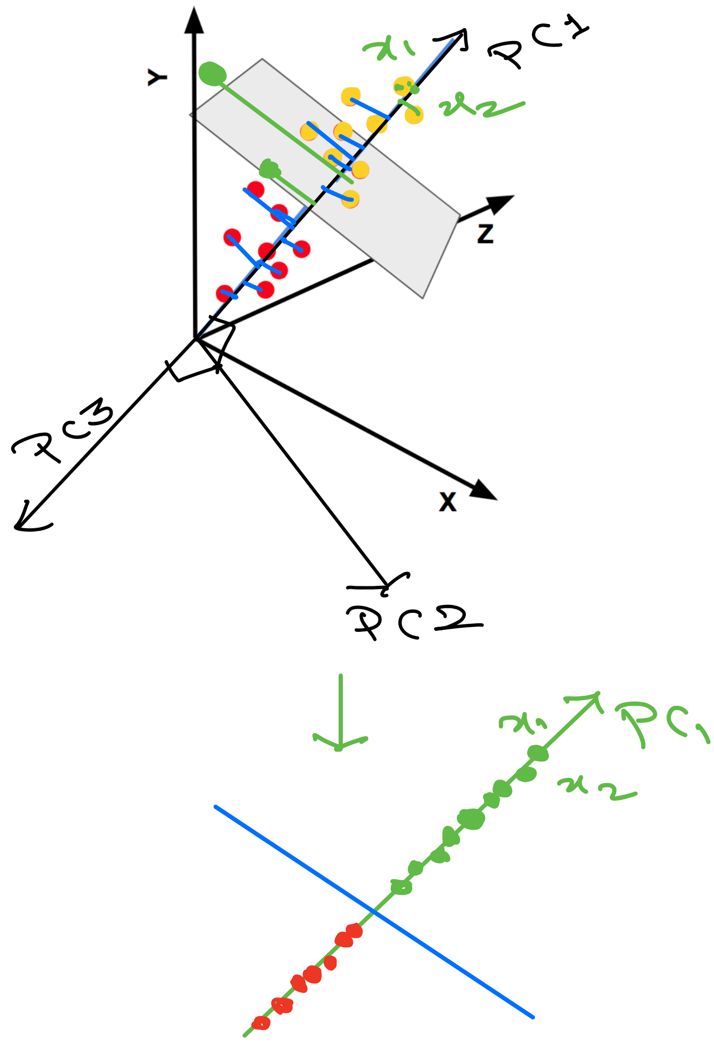


PC1, PC2 and PC3 will orthogonal to each other.

Option 1

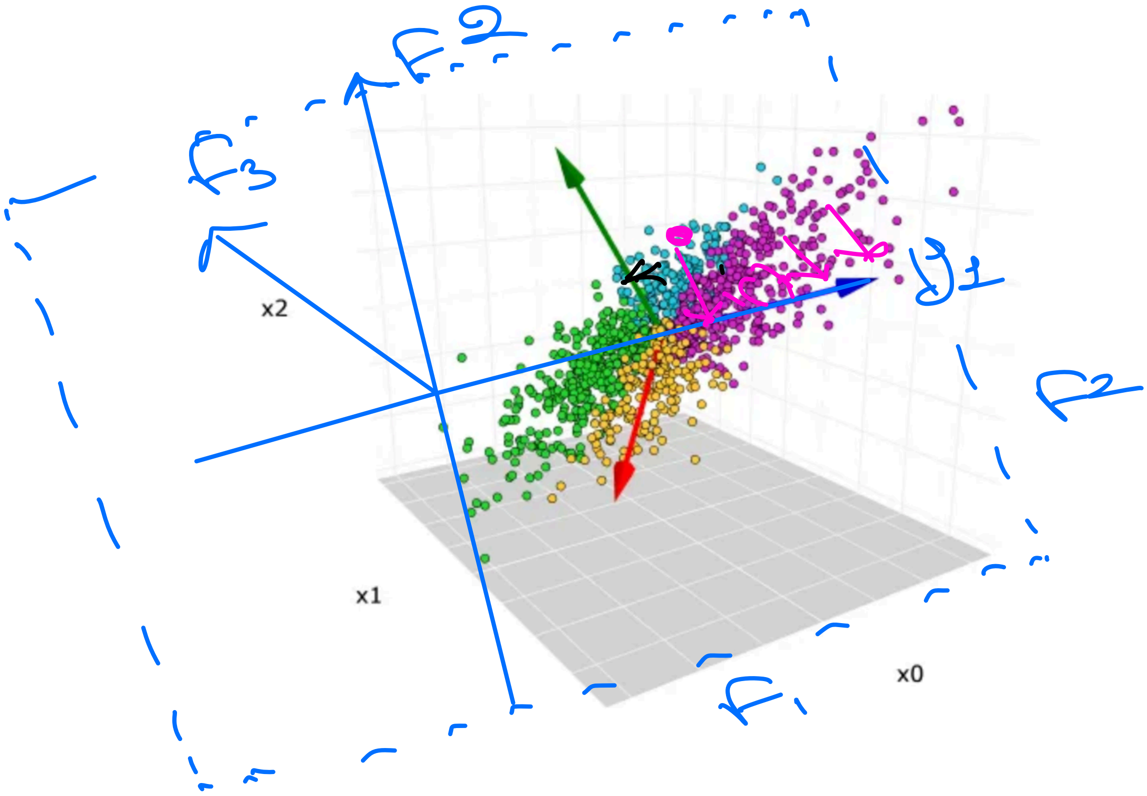


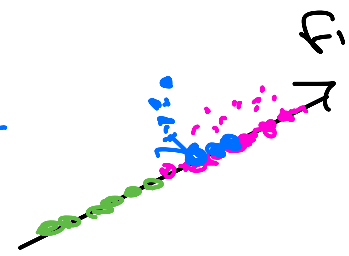
Option 2



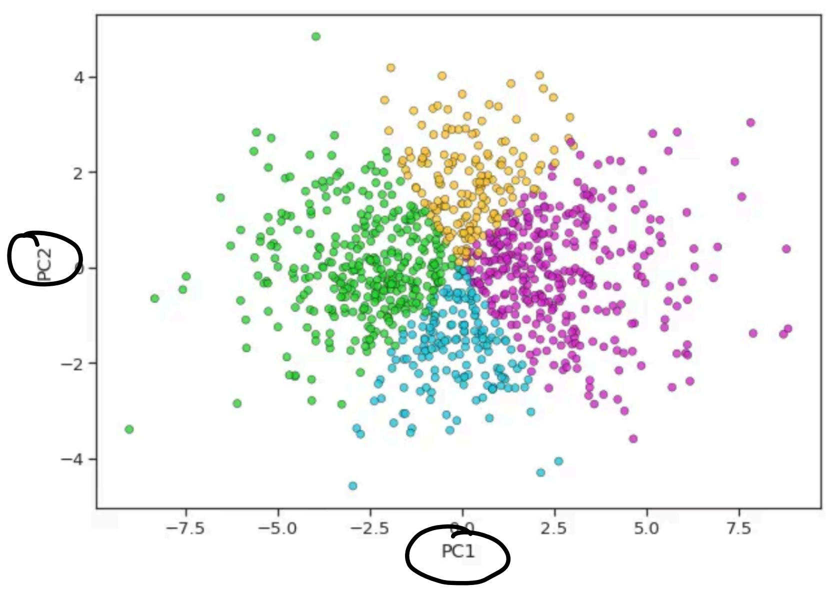
Option 2 is Better Choice

⇒ PC1 with option 2 can give 100% accuracy on classification Even though we lose info stored in PC2 and PC3





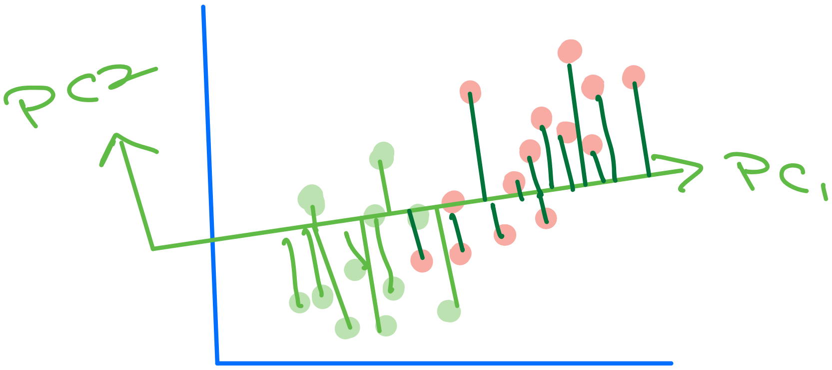
Plotting the D.P. with two PC's



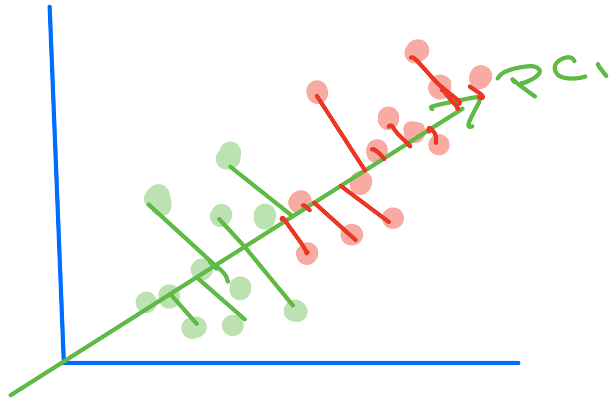
With two PC's we get Non-linear seperable Decision Boundary

⇒ Knn, Decision, RF

Q: Which principal component is better choice



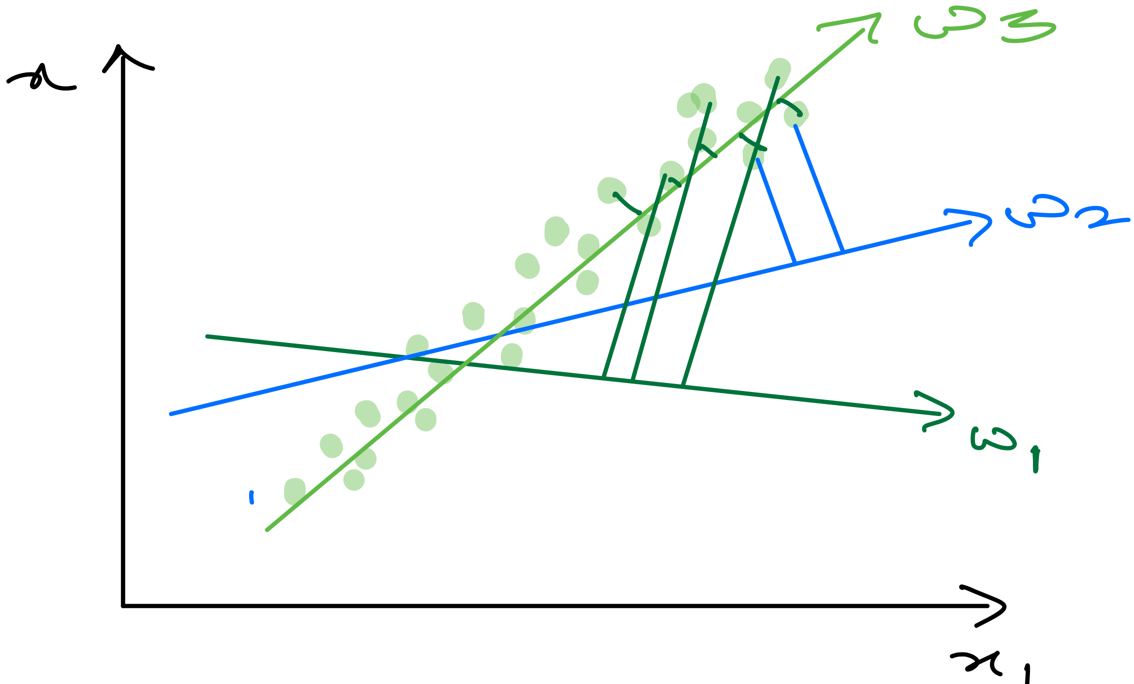
(a)



(b)

Less Loss of Information for this

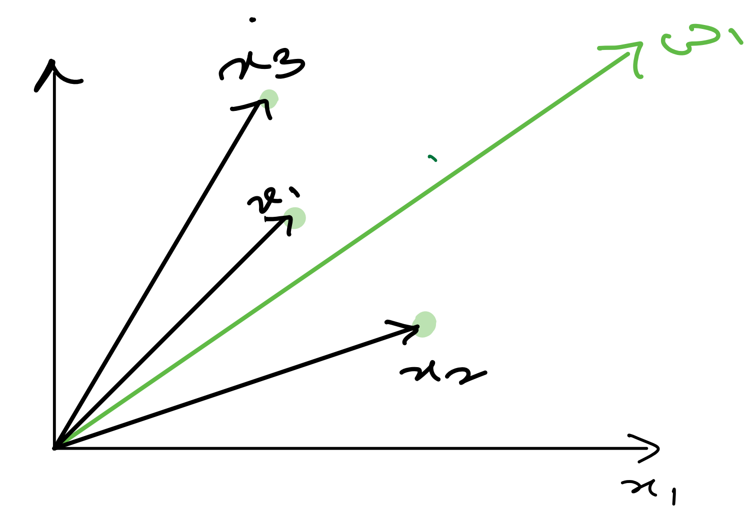
Loss of Info on PC's mathematically

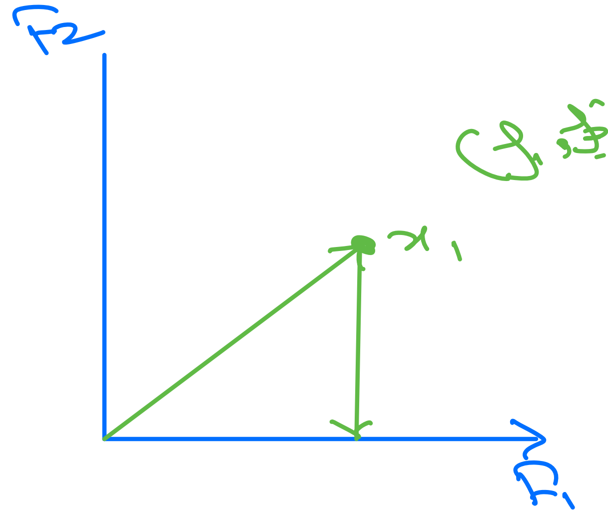


Option-1

Sum of perpendicular distance will be minum for Best PC

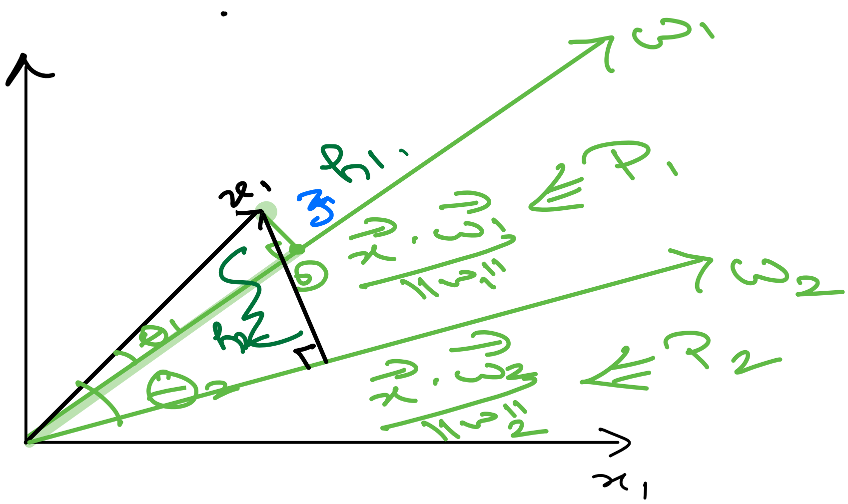
Option 2:

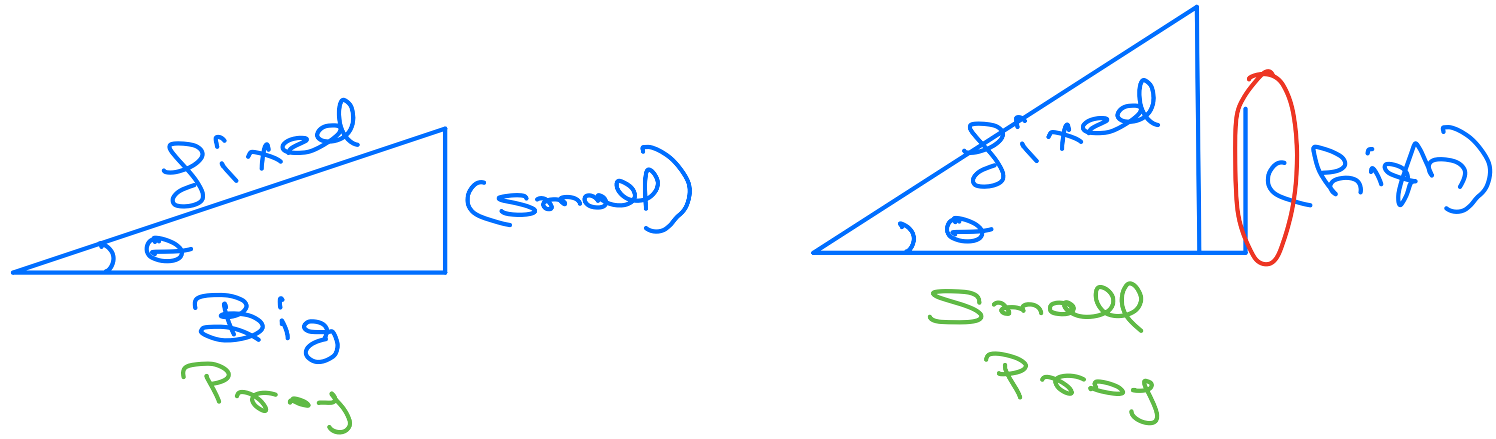




$$\boxed{\text{Proj}_{x \to f_1} \; \frac{\vec{x}_1 \cdot \vec{f}_1}{\|\vec{f}_1\|}}$$

② The Best PC will have maximum projection Length





$$\perp d \propto \frac{1}{Proj}$$

Smaller hieght (Desirable for maximum information) leads to Bigger projection

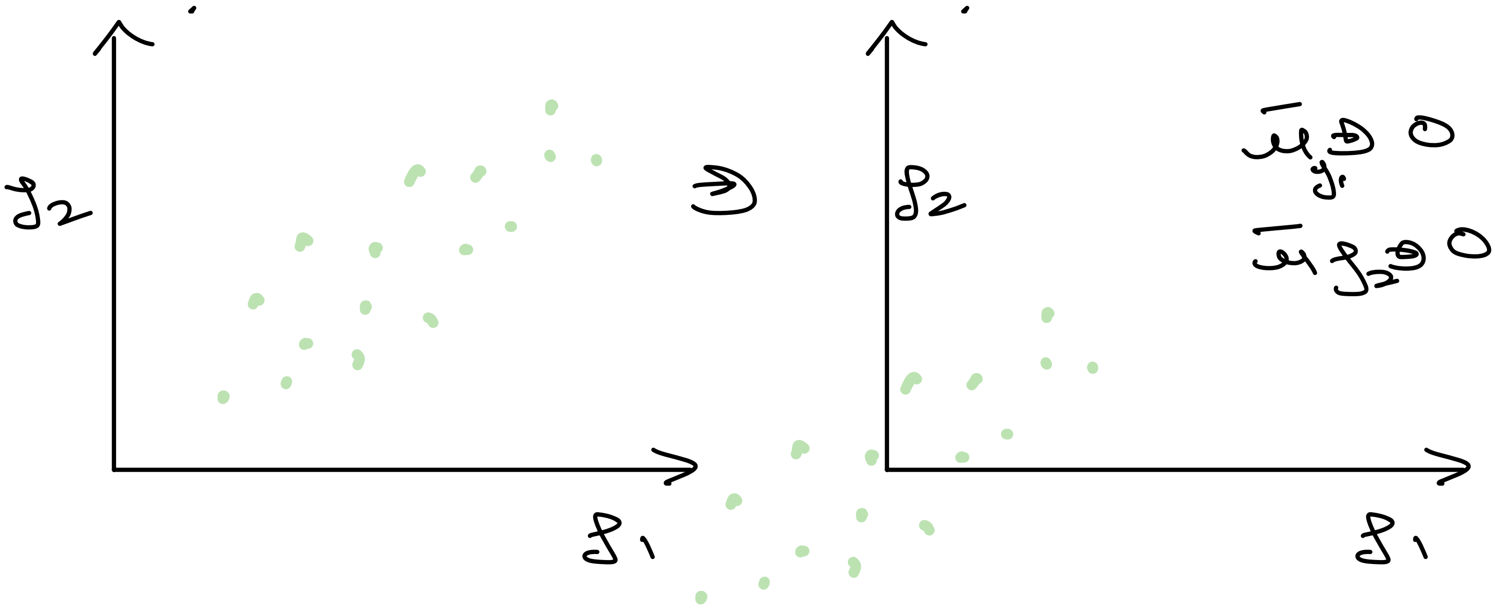
Maths Behind PCA

Step 1: Standardize the data

⇒ PC are robust to outliers

⇒ To avoid dominance of feature with higher Range of values

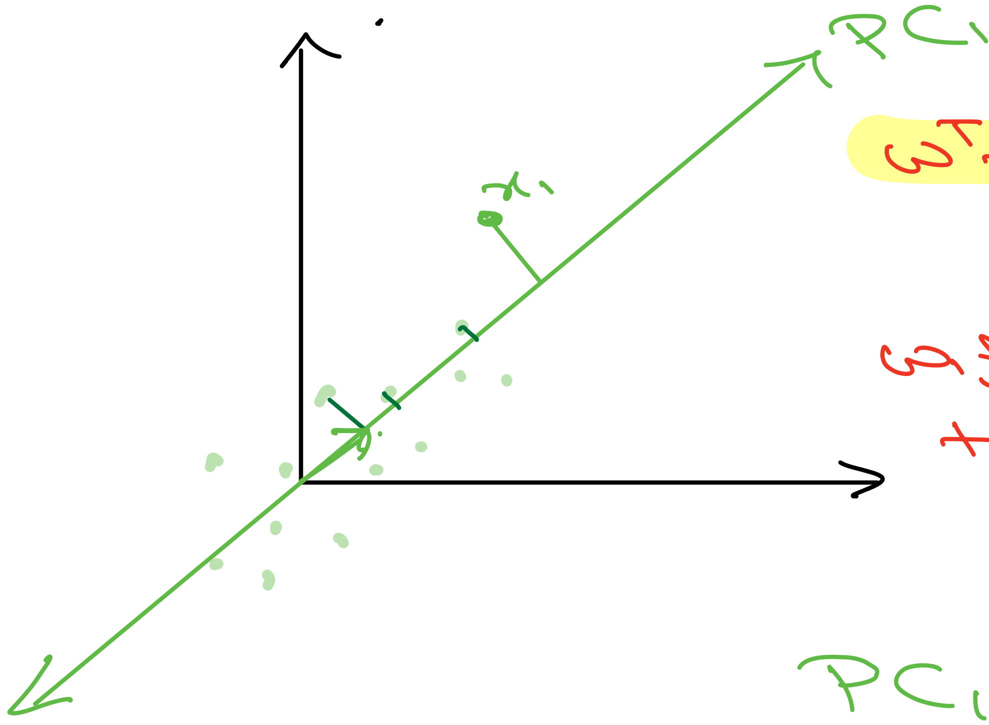
f1 ⇒ −100, 100 } S.D , Mean
f2 ⇒ −1 , 1



$\bar{\mu}_{f_1} \Rightarrow 0$

$\bar{\mu}_{f_2} \Rightarrow 0$

Since Both F_1 and F_2 have mean 0 we assume that PC's will also pass from origin



$w^T x + w_0 = 0$

$\downarrow$

$w_1 f_1 + w_2 f_2 + w_0 = 0$

PC1

$w^T x + \underset{0}{\underset{\downarrow}{w_0}} = 0 \Rightarrow w^T x = 0$

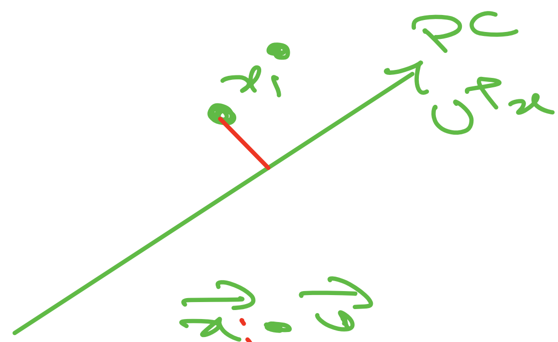
$\downarrow$

$u^T x \Rightarrow 0$

Objective function

Maximize the sum of all projection Lengths

$$\boxed{\max \sum_{i=1}^{n} \frac{\vec{x}_i \cdot \vec{u}}{\|\vec{u}\|}}$$



$$\frac{\vec{x}_i \cdot \vec{u}}{\|\vec{u}\|}$$

Maximize the projection Length

$\Downarrow$

avg Project

$$\boxed{\frac{1}{n} \max \sum_{i=1}^{n} \frac{\vec{x}_i \cdot \vec{u}}{\|\vec{u}\|}} \rightarrow$$

$\vec{u} \Rightarrow$ Unit Vector

$\|\vec{u}\| \Rightarrow 1$

$\Downarrow$

$$\boxed{\max \frac{1}{n} \sum_{i=1}^{n} \vec{x}_i \cdot \vec{u}} \quad \text{s.t. } \|\vec{u}\| = 1$$

Constraint

Constrained optimization

Goal Find Best $\vec{U}$ which maximizes above Equation.

⇒ Solve C.O

⇒ Eigen Values and Eigen Vector

⇒ Implementation of PCA

⇒ T-SNE