# Principal Component Analysis (PCA — A dimensionality reduction technique)

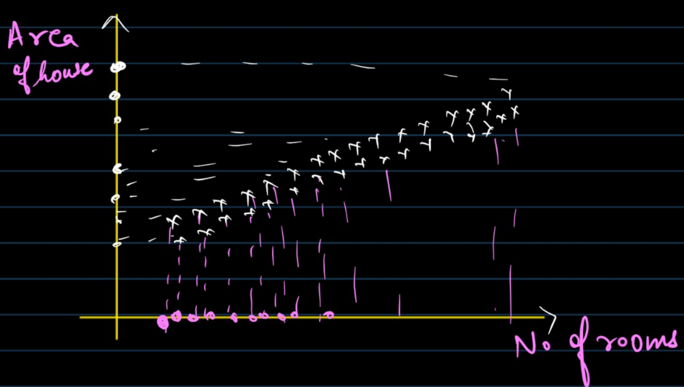
(Geometric Intuition)

| Area of house | No of rooms | Price of house |
|---|---|---|
| | | |

* Since PCA is a dimensionality reduction technique ⇒ We will try to reduce the dimension

* ML is about learning patterns of data.

$\boxed{2d \rightarrow 1d}$ Using PCA.




→ If you want to convert 2d–1d, take only one feature that is number of room or Area of house.

→ Say no of rooms ⇒ x axis you selected.

→ We will loose Area of house.

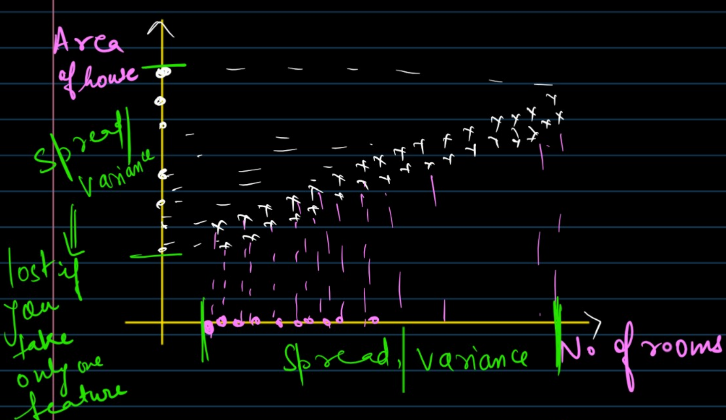
(Actually this also happened in feature selection using correlation)

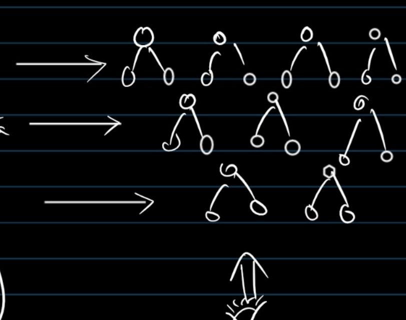
→ But you want both the feature.

→ To get both the features, we need to understand the concept of variance / spread / information.

⇓

Analogy



(from this direction you will click only 3 persons pic)

photographer (captures maximum people in image)

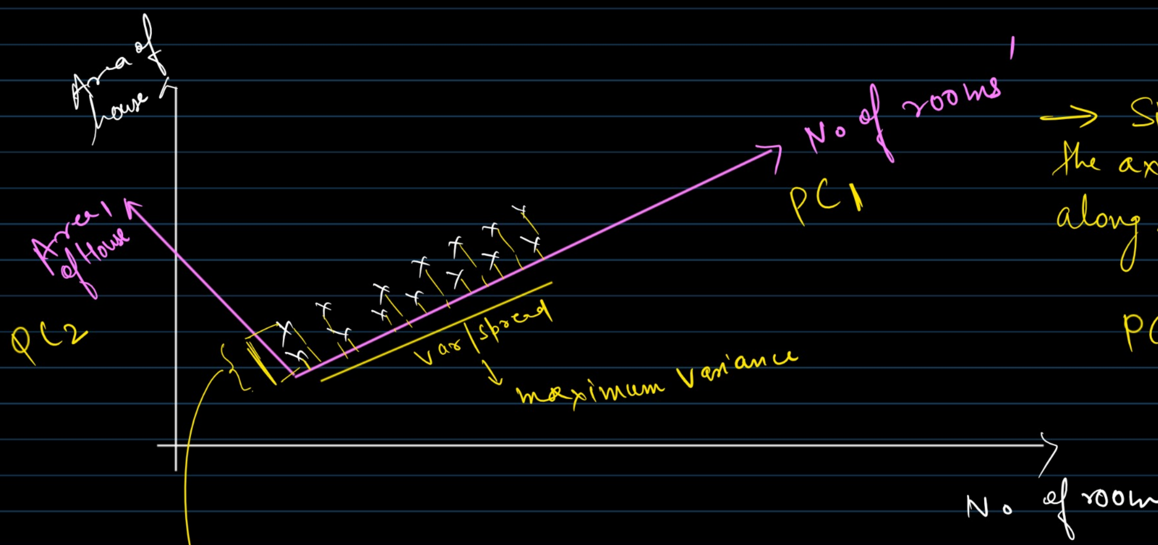
→ If you select only one feature in order to reduce 2d–1d, then spread/variance of other feature is neglected.

Spread ↑ Variance ↑

⇓

maximum spread captured.

→ Spread/variance is pattern of data and ML is about learning patterns, what to do to capture maximum variance and also reduce the dimension.




→ Since you transformed the axis, spread/variance along the axis changed.

⇓

$PC_1$ & $PC_2$

⇓

Still you are getting 2 PC's

2d → $PC_1$ $PC_2$

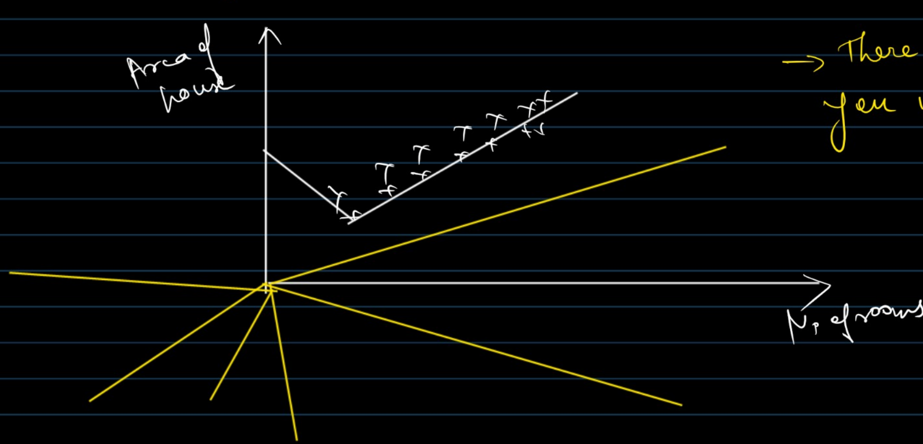
⇓

$PC_1$ will be used for model training as it captures the maximum var/spred of data

→ If you neglect this, less amount of variance will be lost.

2d — 1d ($PC_1$)

→ If you use only new feature No of rooms' ($PC_1$) then very less amount of variance will be lost.




→ There can be n-dimension but you want the maximum variance capture.

| $f_1$ | $f_2$ | $f_3$ | - - - | $f_{1000}$ | $y$ | $PC_1$ | $PC_2$ | $PC_3$ | - - - | $PC_{100}$ |
|---|---|---|---|---|---|---|---|---|---|---|

* No of Principal Components will be equal to number of feature.

* $PC_1 > PC_2 > PC_3 \; \ldots \; PC_n$
  Var Var Var var.

$$f_1 \quad f_2 \quad f_3$$
$$PC_1 \quad PC_2 \quad PC_3$$
$$Var(PC_1) > Var(PC_2) > Var(PC_3)$$

→ first few principal components will capture almost 90% variance/spread/pattern of data. So instead of 1000 PC's you will use only first few.

* In Principal Component $y$ (target variable) is not used, that's why PCA is also called Unsupervised Algorithm.

Axis transformation ⇒ $\boxed{\text{Eigen decomposition of Covariance matrix}}$

* Disadvantage of PCA → loss of information.