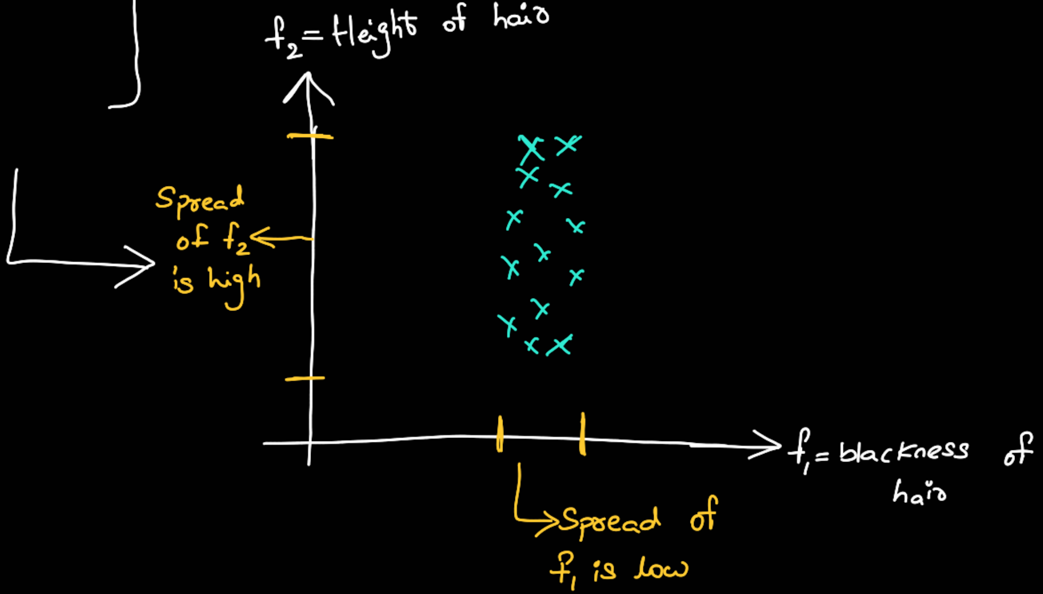
In this lecture lets try to understand the geometric intuition behind PCA. For simplicity lets try reducing from 2D to 1D

Let the features of our data be $\begin{cases} f1 = \text{Blackness of hair of indians} \\ f_2 = \text{Height of hair of indians} \end{cases}$

$$X = \begin{array}{cc} f_1 & f_2 \\ \hline \end{array} \begin{bmatrix} & \\ & \end{bmatrix}$$



So, from the above plot if I am forced to drop one feature I would drop f_1 because the variance along f_1 is low and this will not give much information. So, we would drop a low variance feature by which we will retain most of the information

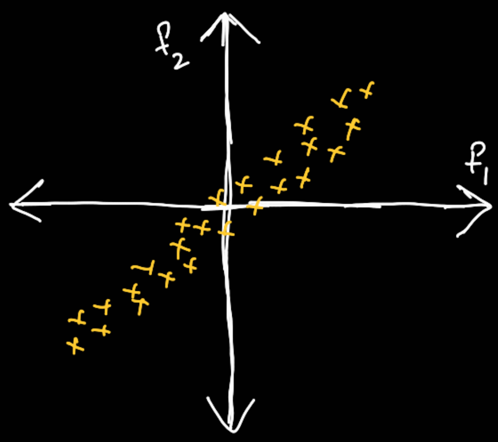
$$X = \begin{array}{cc} f_1 & f_2 \end{array} \begin{bmatrix} & \\ & \end{bmatrix} \xrightarrow{\text{dimensionality reduction}} X' = \begin{array}{c} f_2 \end{array} \begin{bmatrix} \\ \end{bmatrix}$$

Preserving the direction with maximal spread/variance

But this is always not so simple. Conider

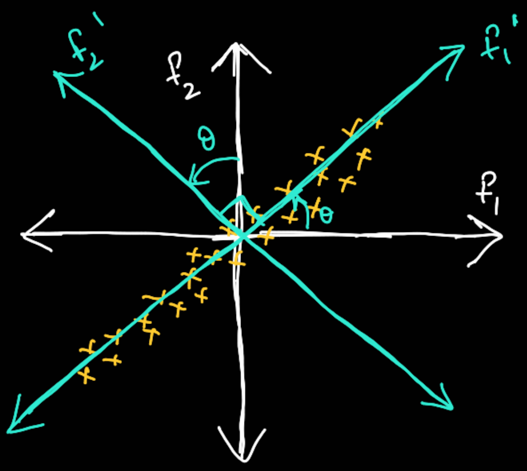
X = 2-dim dataset

Col-standardized $\begin{cases} mean\{f_1\} = mean\{f_2\} = 0 \\ var\{f_1\} = var\{f_2\} = 1 \end{cases}$



Here the spread is not just on f_1 or f_2. you can't simply drop a axis which will result in more information loss.

So, here is what we can do.



① Find $f_1' \perp f_2'$ s.t f_1' has max spread
(spread of f_2' << spread of f_1')

② Drop f_2'

③ Project x_i's onto f_1'

So, here we are rotating our axis to find f_1' with max-variance and then drop f_2'

In the next lecture lets see how to do this mathematically.