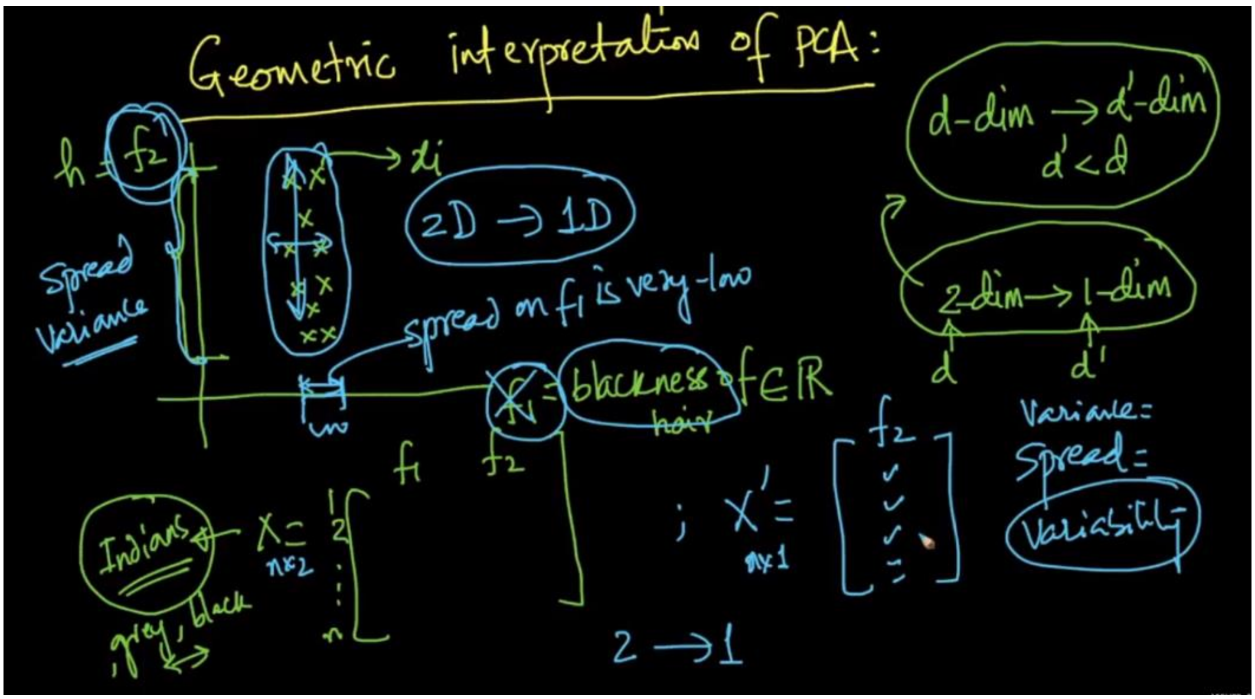
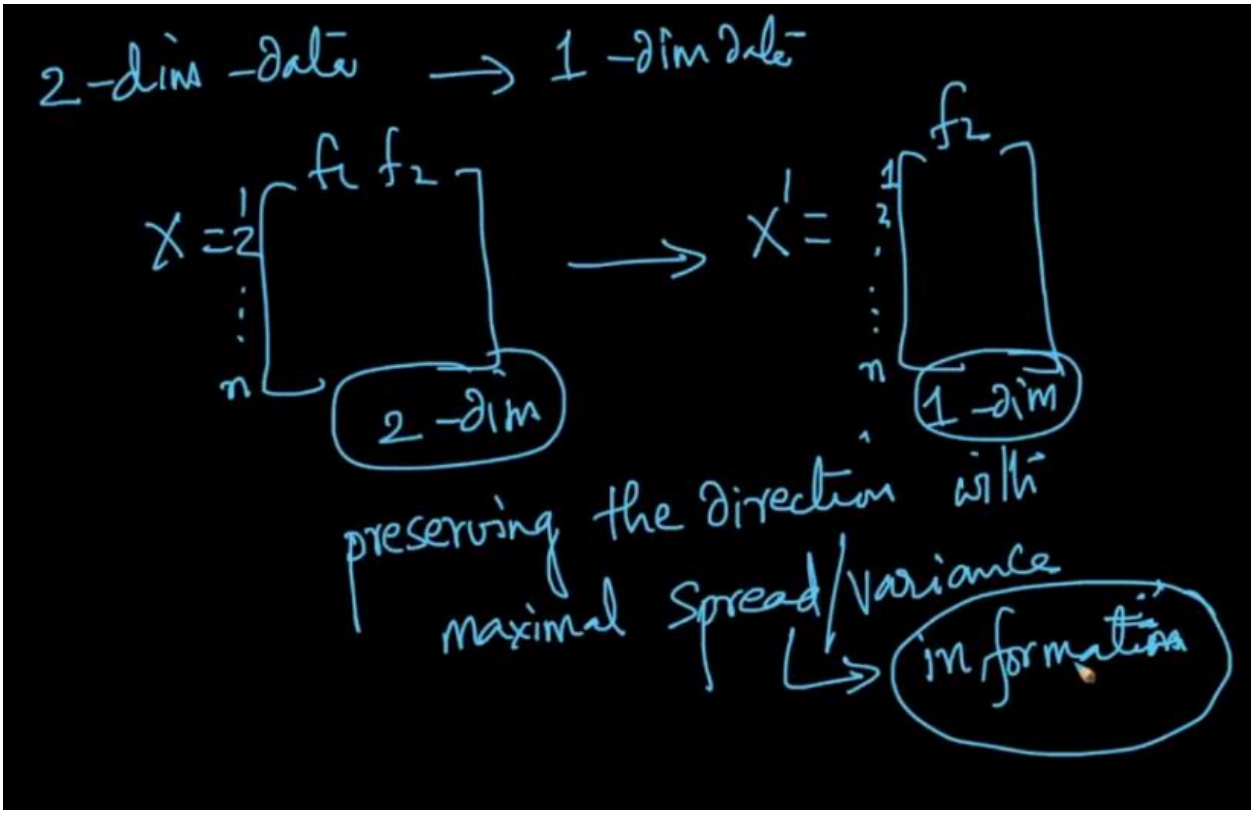
Geometric Interpretation of PCA:

Since in PCA we convert high dimensions into comparatively low dimensions, so for understanding let's say we convert 2-D to 1-D.

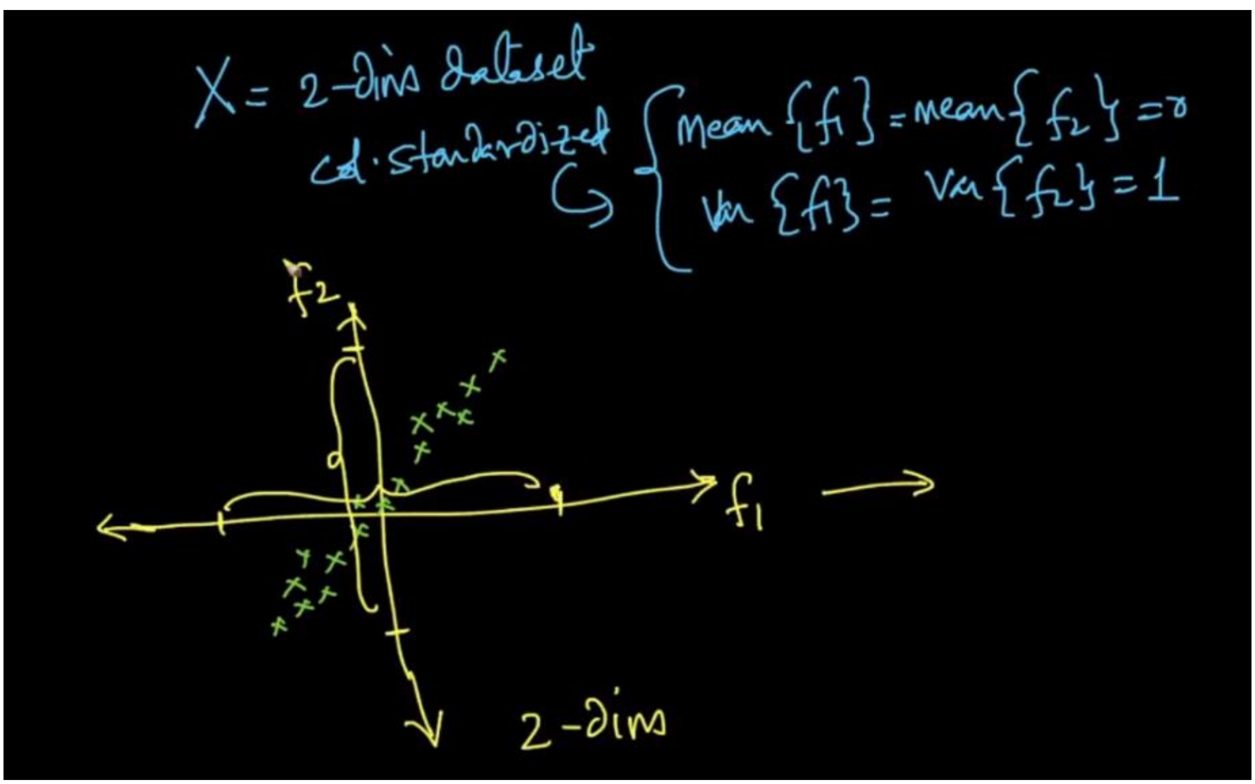
Given a data of Indians with their height(f2) and blackness of hair(f1), now after plotting them, we can say that spread or variance is more on height(f2) as compare to f1, so if we are compelled to convert 2-D to 1-D then we remove f1. And take f2 as our final feature.

We always prefer keeping the feature with more spread/variance, because it gives much more information.



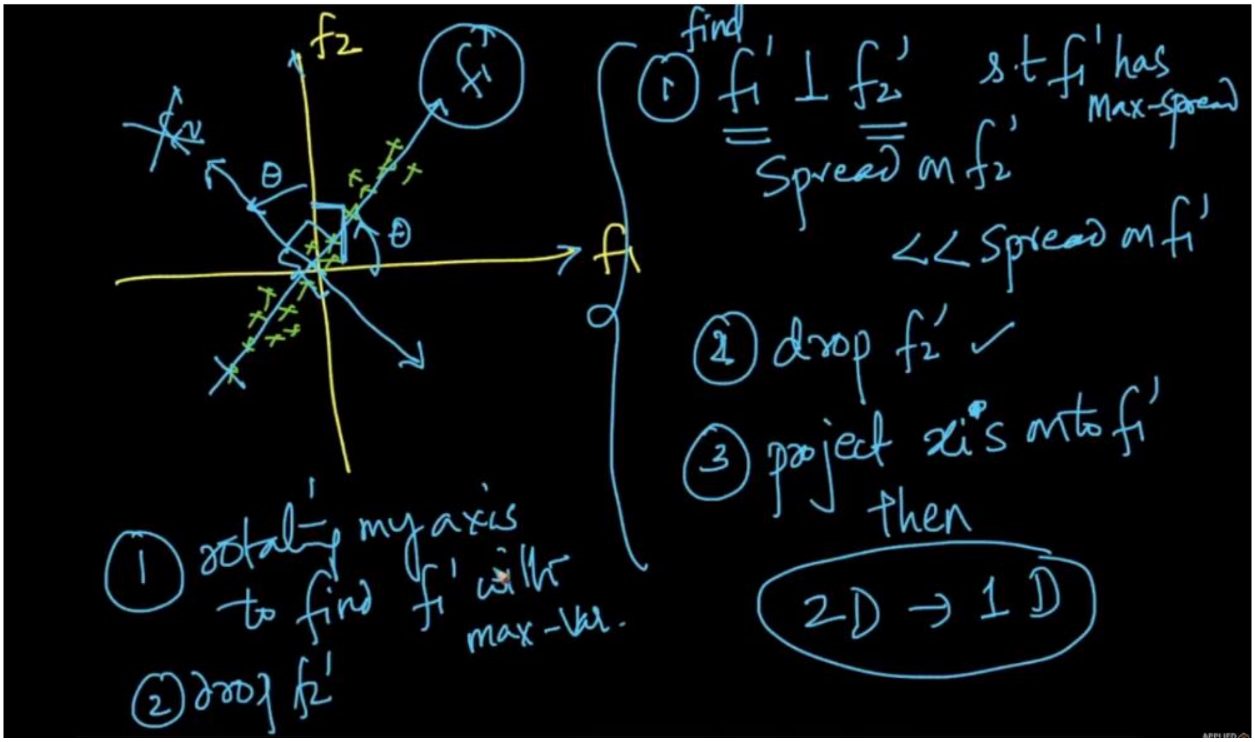


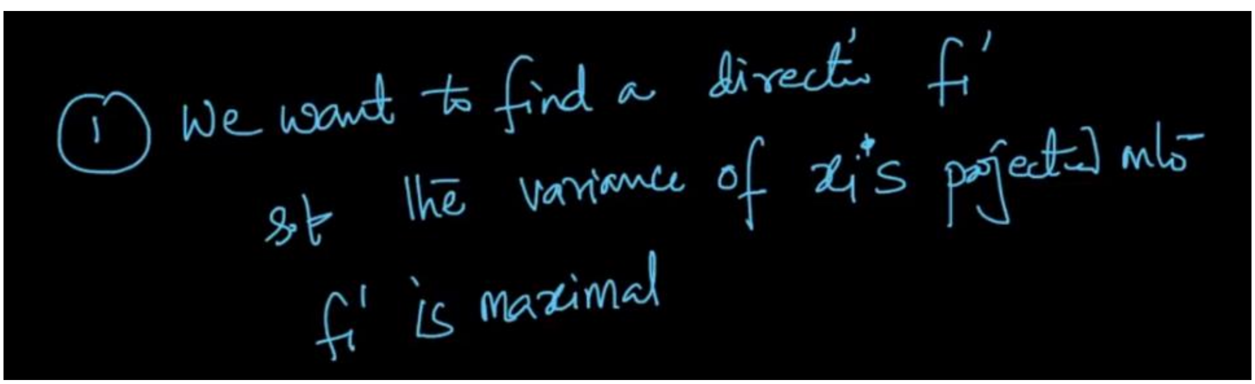
Now let's understand with new example where the given dataset is column standardized, now in this case both the features f1 and f2 have variance of 1, so now to decide which feature to remove is difficult.



So in such case we will find the new direction (by rotating existing features) on which variance of given x_i's will be maximum, let's say this f1', we also find a direction perpendicular to f1', let's say this as f2'.

Since spread is much more on f1' as compare to f2' and therefore we keep f1' and remove f2', and in this way we convert 2-D to 1-D.





Notes:

- if features are highly correlated then should i apply PCA?
  if features are highly(here highly means extremely high) correlated then it is better to remove them.there will be some effect on your result with this.

- what if features are completely uncorrelated?
  PCA is useless for dimensionality reduction if features are completely uncorrelated
- Should we check features are correlated or not before applying PCA?
  you can check correlation and if some features are extremely highly correlated then you can remove them.