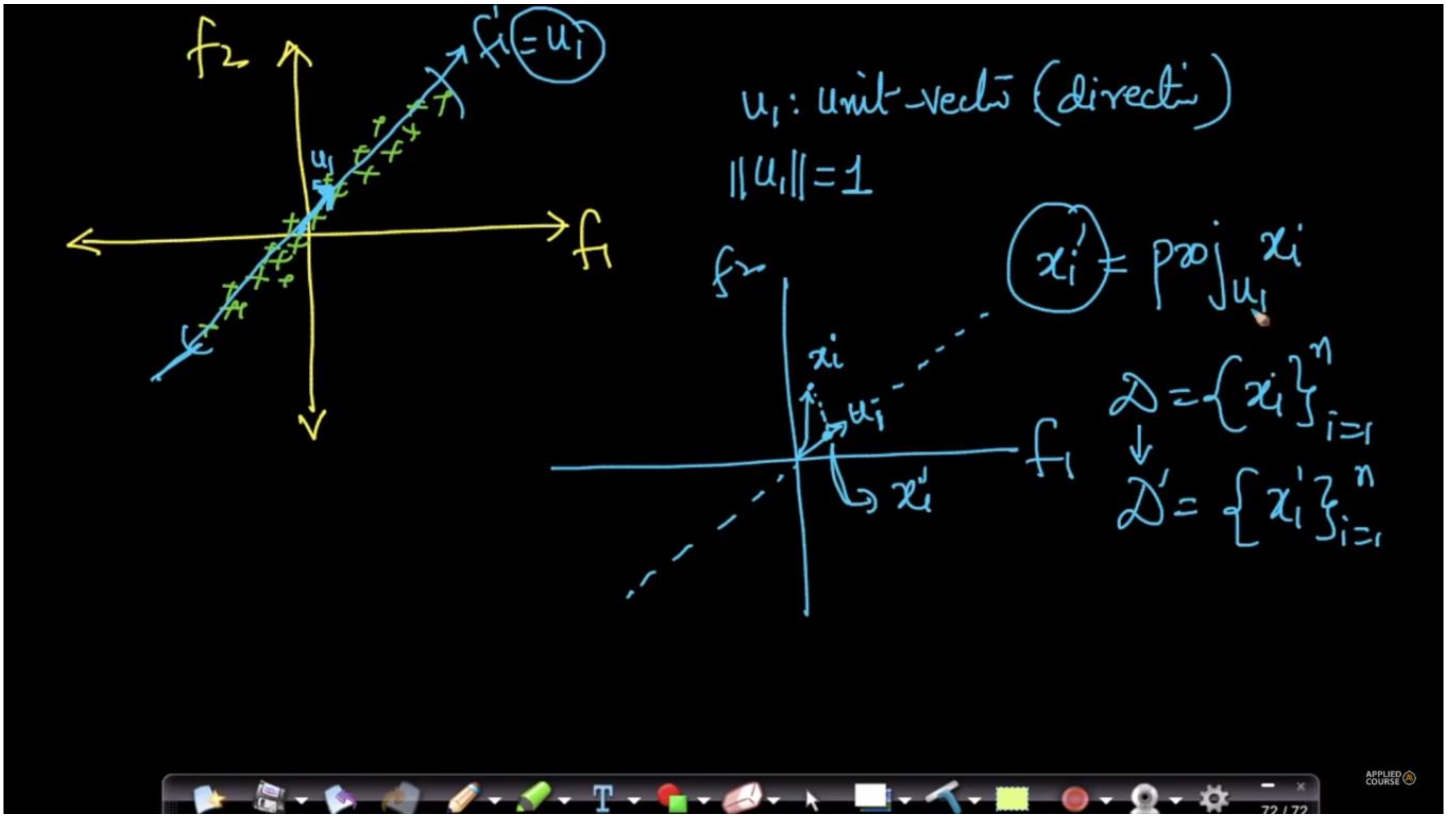
In this we see the math behind finding the new direction where the variance of x_i’s will be maximum.

Basically what are we doing is picking up a unit vector (u1) and projecting x_i on u1.

so we’ll get new x’s such that $x_i' = proj_{u1}\ x_i$. and hence we get new dataset as $D' = \{ x_i' \}_{i=1}^{n}$

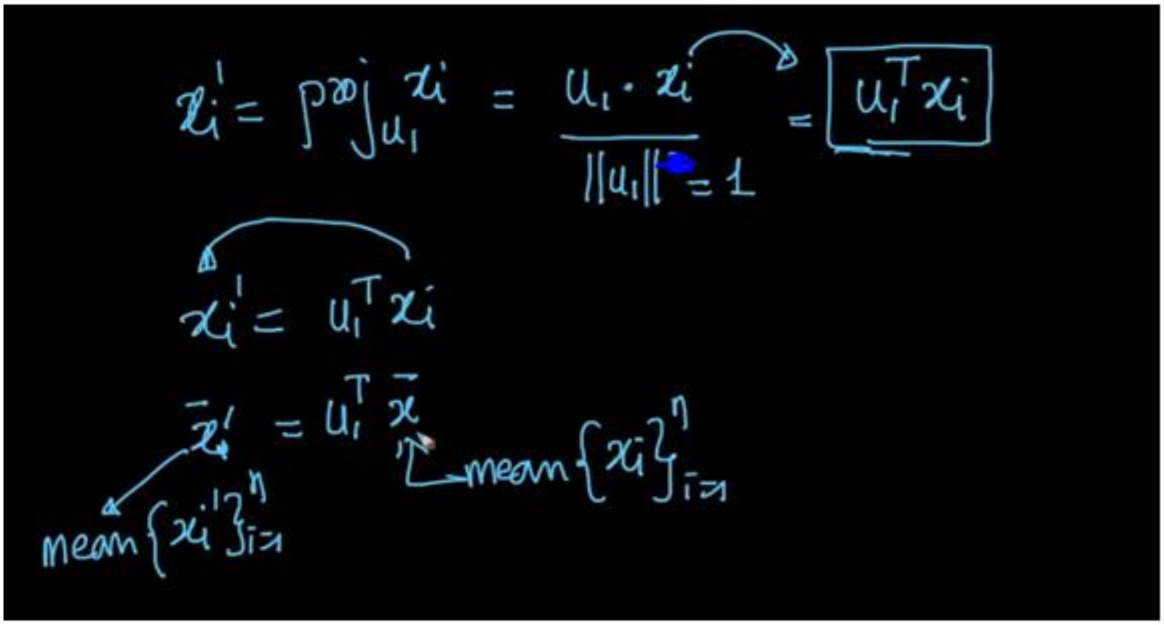


Since we know that project of any point (x_i) on a vector (u_i) is given as mentioned in below fig.

And since unit vector’s value is 1, therefore x_i' will be:

$$\mathbf{x_i' = u_1^T \, . \, x_i}$$

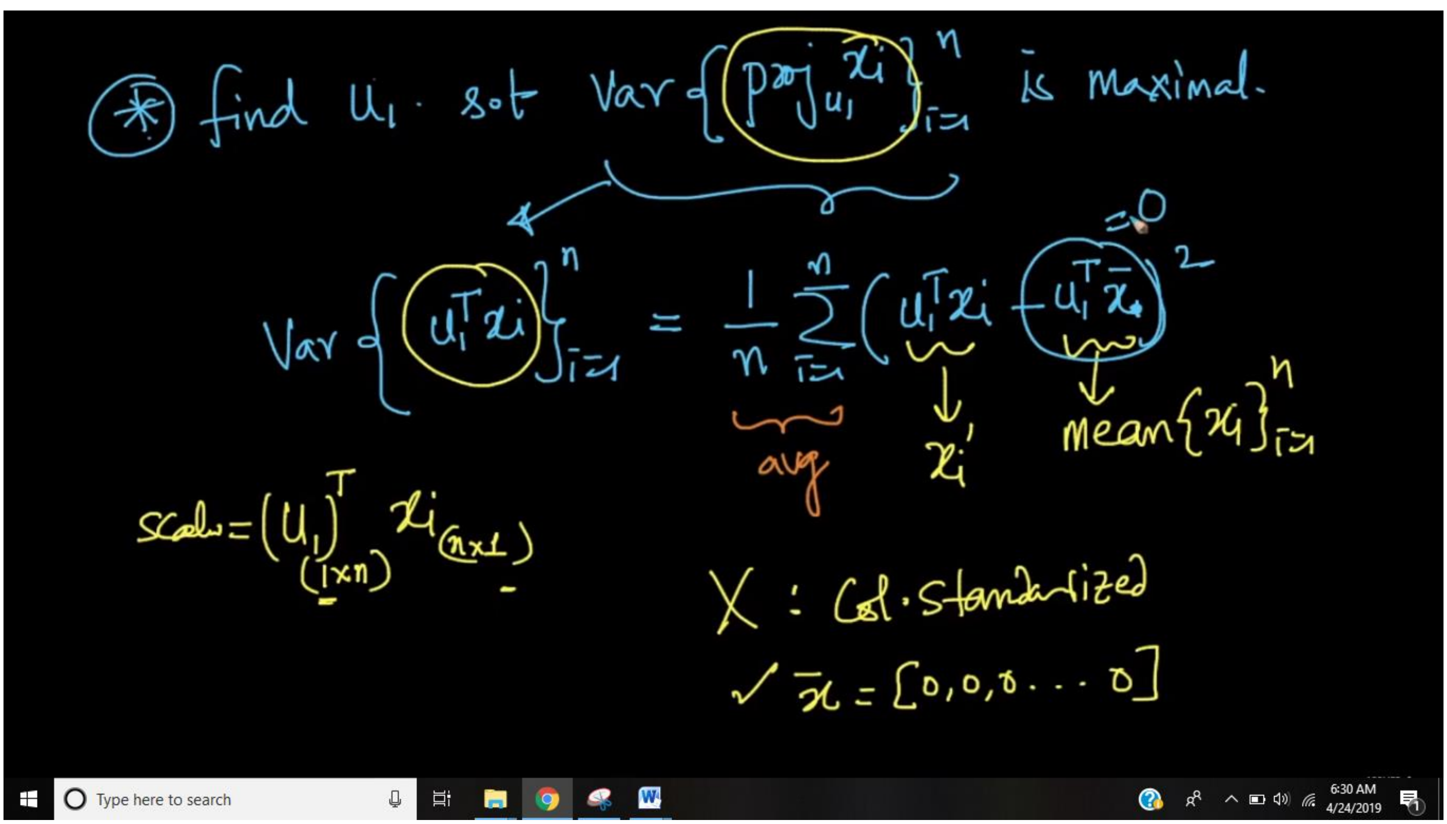
Now same we do with mean of x_i’s mean, that means we project x_i s mean on u1 and we get mean of x_i'



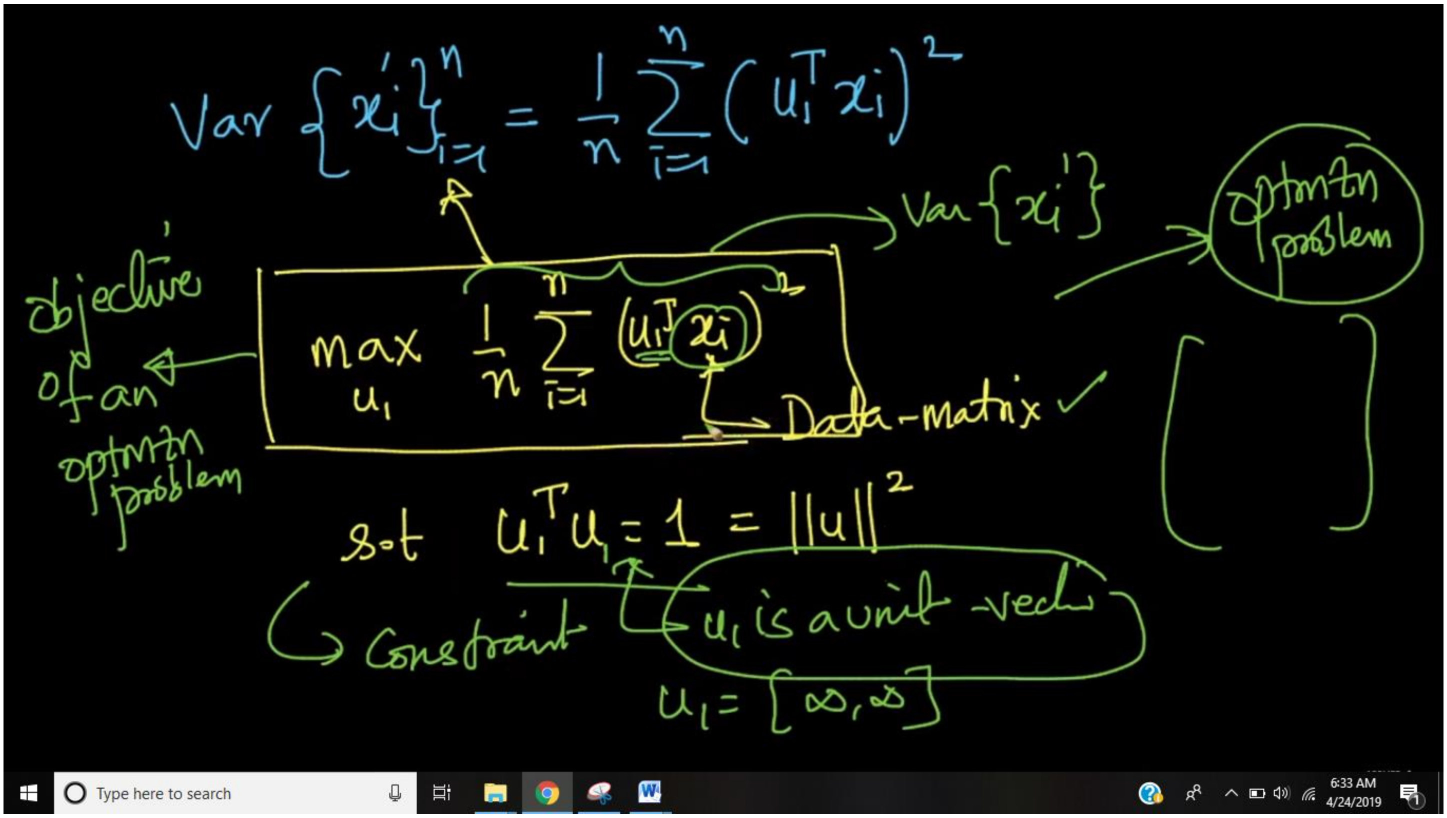
Now our main objective is to find the direction where variance (variance of x_i projected on u_1) is maximal.

And by using variance formula we can find that, but if our dataset is column standardized, then we remove mean term from that because mean will be 0. Given in below fig.

One thing to note here is that u_1^T is of dimension 1 * n and x_i is of dimension n * 1 and therefore there multiplication will be scalar.



So our ultimate aim is find the direction u1 such that Variance will be maximum in a given constraint that $u^T u = 1$.



Summary:

If we have a 2D space.
Feature1 = weight (lets say)
Feature2 = height (lets say)

As a first step we would want to apply standardization for each feature first.
After applying the transformation we have a 2D data set centered as the origin with std of 1.
This implies that each of our variables would have the exact same spread (and therefore we would not know at this point which feature is the better one to use)

We will therefore always need to do an axis transformation (find unit vector) in order for the axis to lineup with features that have the highest spread?

* the unit vector = just a new feature created (with its data observations the projected values) ?