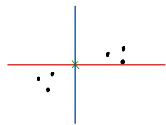


Principal Component Analysis

original 2D data
plot by dimension



center on origin

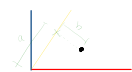


find line of best fit

rotate line about origin

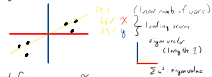


how do we know when line is "best"?



maximize E_1^2 or minimize E_2^2
variance math

We get:

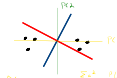


Repeat for remaining PCs

PC1 is the direction of max variance
PCs must always be \perp to PC_{i-1}



rotate to make final PCA plot



PC1: direction of E_1^2 (largest eigenvalue)
PC2: direction of E_2^2 (2nd largest eigenvalue)

ex: PC1 var = 15
PC2 var = 5

PC1 accounts for $\frac{15}{20} = 75\%$ of total variation
PC2 accounts for $\frac{5}{20} = 25\%$ of total variation

Scree Plot:

