Principal Component Analysis(PCA) problem formulation

e.g.: Reduce from 2-dimension to 1-dimension: Find a direction(a vector u\_1 є Rn) onto which to project the data, so as to minimize the projection error.

Generally, Reduce from n-dimension to k-dimension: Find k vectors u\_1, u\_2, u\_3, …, u\_k onto which to project the data, so as to minimize the projection error.

Before applying PCA, there is a data preprocessing step.

Training set: x\_1, x\_2, x\_3, …, x\_m.

Preprocessing(feature scaling/mean normalization):

μj = 1/m \* sum(xij)(i from 1 up to m)

replace each xij with xj – uj

If different features on different scale(e.g., x1 = size of house, x2 = number of bedrooms), scale features to have a comparable range of values.