1. (Math) Consider how to find the first principal component v for data points $x_1, x_2, \cdots, x_n$. One can maximize the variance of the projections (suppose the data have mean $\frac{1}{n}\sum_{i=1}^{N} x_i = 0$):

$$\max_{v:\|v\|_2=1} \sum_{i=1}^{n} (v^T x_i)^2$$

One can also minimize the mean square error of the projections:

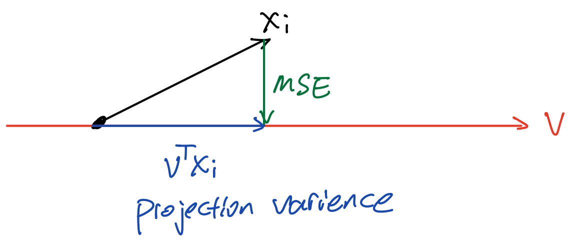
$$\min_{v:\|v\|_2=1} \frac{1}{n} \sum_{i=1}^{n} \|x_i - vv^T x_i\|_2{}^2$$

Show that the optimal solution to the two is the same.

Hint: see the intuition on slide 18 of "Deep Learning Lecture 7: Factor Analysis".

Prove that maximize the variance of the projection is equivalent to minimize the mean square error.

① Pathagorean Theorem



X_i is the data, fixed

V is the direction

Mean Square Errow is the vector othogonal to the direction V

$\frac{1}{n}(x_i - vv^T x_i)^2 = MSE$

Projection Varience is the vector projected by data X_i with direction V.

$V^T x_i$ = Magnitude of Project Varience.

$$[V^T x_i]^2 + [MSE]^2 = [X_i]^2$$

From the Pathagorean Theorem, we can conclude that maximize $V^T x_i$ is equivalent to minimize MSE

If we minimize MSE to 0.

X_i will equal to it's projection on V.

In this case, the optimal solution is achieved

② First Principal Component

Definition: PC1 is the line that best accounts for the shape of the points.

PC1 represents the direction that the data spread the most.

If the data are all linearly spread on this direction, there will be no mean square error which measures how much the data are othogonally away from the direction.

Conclusion, PC1 and projection yeild MSE are othogonal to each other and inorder to find the optimal solution, we should maximize the PC1 and minimize the yeild.