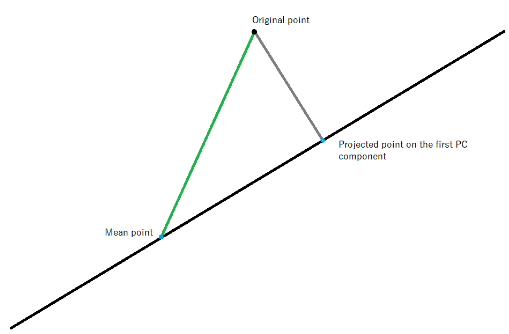
Authors: Mhd Jawad Al Rahwanji, Wojciech Ptaś 7038980, 7042843

Problem 1

In order to proof that the first PCA minimizes the residual sum of squares I will use a graphical representation. Here, the RSS is represented by the length of the gray line between original data point and its projection. The distance between the mean point (i.e. 0, because the data should be zero centered) and the projected data point is variance of this data point. The distance between the original data point and the mean is a constant value.



This three lines create a right triangle. By Pythagorean theorem it means that:

$(\mu - x_{ij})^2 = (\mu - \hat{x}_{ij})^2 + (x_{ij} - \hat{x}_{ij})^2$

where μ is the mean, and $\hat{x}_i$ is the data point projected onto the first PCA component.

We can now sum this equation over all data points:

$\sum_i (\mu - x_{ij})^2 = \sum_i (\mu - \hat{x}_{ij})^2 + \sum_i (x_{ij} - \hat{x}_{ij})^2$

Now if we divide both sides by number of data points, we get

$\frac{1}{n} \sum_i (\mu - x_{ij})^2 = \frac{1}{n} \sum_i (\mu - \hat{x}_{ij})^2 + \frac{1}{n} \sum_i (x_{ij} - \hat{x}_{ij})^2$

Which can be rewritten as:

$const = Var_{z_1}(x) + \frac{1}{n} RSS$

This equation shows, that if the PCA maximizes the variance caputed by first principial component it also has to minimize the RSS becaause this two values add up to a constant.