11/11/2021

In-Class Problem 12

Prove that for PCA : $\sum_{i=1}^{N} x_i = 0$

⇒ Let's start this problem by taking a note of the fact that in PCA, we centralize the dataset first by subtracting it's mean.

So, $\bar{x} = x_i - \mu$ where x is the data points and μ is the mean

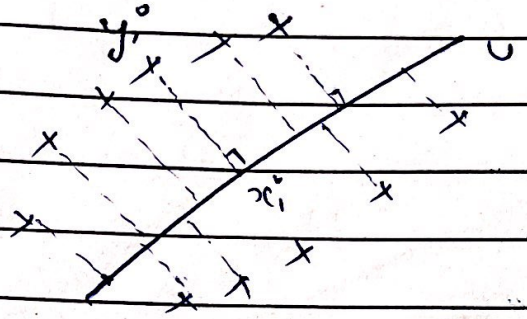
Now, we define a PCA optimzation problem as,

$$\begin{cases} \min \| y_i - (v x_i + \mu) \|_2^2 \\ \text{s.t.} \quad v^T v = 1 \\ \sum x_i = 0 \end{cases}$$

Here, v is the principal component $[v \in R^{d \times p}]$ and μ is the offset of the projection from origin (not the mean).

Date ___/___/___
Page ___

Therefore → $\|y_i - (wx_i + u)\|_2^2$ is the distance of the original ~~point~~ point from it's projection on the principal component.



As we can observe here, $\sum x_i$ here is nothing but the mean of all the projections which is equivalent to the mean of all y_i's.

And we know that the mean of the data is 0 since we centralized all the points.

$$\therefore \sum x_i = 0$$