✍2 Show that $\bar{X} = X(I - \frac{1}{n}ee^T)$ (here e = vector of all ones). What does the projector $(I - \frac{1}{n}ee^T)$ do?

Solution: Each column of $\bar{X}$ is $\bar{x} = x - \mu$ so that $\bar{X} = X - \mu e^T$, where μ is the sample mean. But we have $\mu = \frac{1}{n}\sum x_i = \frac{1}{n}Xe$ and so,

$$\bar{X} = X - \frac{1}{n}Xee^T = X[I - \frac{1}{n}ee^T]$$

The matrix $(I - \frac{1}{n}ee^T)$ represents a projector that centers the data so the mean is zero.

✍3 Show that solution V also minimizes 'reconstruction error' ..

Solution: The main property that is exploited in the proof is the fact that $\text{Tr}(ABC) = \text{Tr}(BCA)$ (when dimensions are compatible). First we note that $\sum_i \|\bar{x}_i - VV^T\bar{x}_i\|^2 = \|(I - VV^T)X\_F^2$. We will call

$\boldsymbol{P}$ the pojector $\boldsymbol{P} = \boldsymbol{V}\boldsymbol{V}^T$. Then:

$$
\begin{aligned}
\|(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^T)\boldsymbol{X}\_\boldsymbol{F}^2* &= \operatorname{Tr}(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{X}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{P}) \\
&= \operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^T - \boldsymbol{P}\boldsymbol{X}\boldsymbol{X}^T)(\boldsymbol{I} - \boldsymbol{P}) \\
&= \operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^T) - \operatorname{Tr}(\boldsymbol{P}\boldsymbol{X}\boldsymbol{X}^T) - \operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{P}) + \operatorname{Tr}(\boldsymbol{P}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{P}) \\
&= \operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^T) - \operatorname{Tr}(\boldsymbol{P}\boldsymbol{X}\boldsymbol{X}^T) - \operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{P}) + \operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{P}^2) \\
&= \operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^T) - \operatorname{Tr}(\boldsymbol{P}\boldsymbol{X}\boldsymbol{X}^T) - \operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{P}) + \operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{P}) \\
&= \operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^T) - \operatorname{Tr}(\boldsymbol{P}\boldsymbol{X}\boldsymbol{X}^T) \\
&= \operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^T) - \operatorname{Tr}(\boldsymbol{V}\boldsymbol{V}^T\boldsymbol{X}\boldsymbol{X}^T) \\
&= \operatorname{Tr}(\boldsymbol{X}\boldsymbol{X}^T) - \operatorname{Tr}(\boldsymbol{V}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{V})
\end{aligned}
$$

The first term is a constant, therefore the minimum is reached when the maxiumn of the second term is reached. $\square$

✍4 ... and that it also maximizes $\sum_{i,j} \|\boldsymbol{y}_i - \boldsymbol{y}_j\|_2^2$

Solution: Let us denote by $\bar{\boldsymbol{y}}$ the sample mean of the $\boldsymbol{j}_j$s, i.e.,

$$\bar{\boldsymbol{y}} = \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{y}_j.$$

We proceed backward examine the sum $\sum_{i,j} \|y_i - y_j\|_2^2$

$$\begin{aligned}
\sum_{i,j} \|y_i - y_j\|_2^2 &= \sum_{i,j} \|(y_i - \bar{y}) - (y_j - \bar{y})\|_2^2 \\
&= \sum_{i,j} ((y_i - \bar{y}) - (y_j - \bar{y}), (y_i - \bar{y}) - (y_j - \bar{y})) \\
&= \sum_i \sum_j \left[\|(y_i - \bar{y})\|_2^2 + \|(y_j - \bar{y})\|_2^2\right] \dots \\
&\quad - 2 \sum_i \sum_j ((y_i - \bar{y}), (y_j - \bar{y})) \\
&= 2n \sum_i \|y_i - \bar{y}\|_2^2 - 2 \sum_i (y_i - \bar{y}, \sum_j (\bar{y} - y_j)) \\
&= 2n \sum_i \|y_i - \bar{y}\|_2^2
\end{aligned}$$

The last equality comes from the fact that $\sum_j (\bar{y} - y_j) = 0$ $\square$

Further reading: Some references on applications of the SVD

For image processing:

https://arxiv.org/pdf/1211.7102.pdf

The extraordinary SVD

https://arxiv.org/pdf/1103.2338.pdf%22%20rel=%22nofollow

An outstanding paper for understanding the SVD and a few of its applications:

https://sites.math.washington.edu/~morrow/464_16/svd.pdf

An old paper of ours that discusses a form of truncated SVD for face recognition:

https://www-users.cs.umn.edu/~saad/PDF/umsi-2006-16.pdf