02/02/2023 PCA Continued

$B = \begin{bmatrix} - b_1 - \\ - b_2 - \\ \vdots \\ - b_n - \end{bmatrix}_{400 \times 4096}$ centered face dataset

$\bar{a}$: mean face image

$B = A - \bar{a}$

↑ original face dataset

1) $B = U \Sigma V^T$

U: 400x400, Σ: 400x4096, V^T: 4096x4096

2) To get d dimensional representation of B ($d << 4096$)

$T_B = BV[:, [0, 1, \ldots, d-1]]$

T_B: 400xd, B: 400x4096, $V[:, [0, 1, \ldots, d-1]]$: 4096 x d

★★ Principal Component is the best linear approximation in lower dimensionality (d) ★★

3) To get d dimensional representation of A (T_A)

$T_A = T_B + \bar{a} \cdot V[:, [0, 1, \ldots, d-1]]$

T_A: 400xd, T_B: 400xd, $\bar{a} \cdot V[:, [0, 1, \ldots, d-1]]$: 1xd

- Where is the most information in SVD?
  - if B has columns for samples, then U columns
  - if B has rows for samples, then V^T row
- Variance
  1) Let A be a matrix whose columns contain separate samples
  2) $A = U \Sigma V^T$ (SVD)
  3) If you take first j columns of U variance of the data A that you will capture:

$$\frac{\sum_{i=1}^{j} \sigma_i^2}{\sum_{i=1}^{\min(m,n)} \sigma_i^2}$$

j = rank of A

variance captured = 1

$j <$ rank of A

$0 <$ variance captured < 0