COM S 5730 Bonus Homework

1. Please put required code files and report into a compressed file "BHW_FirstName_LastName.zip"
2. Unlimited number of submissions are allowed on Canvas and the latest one will be graded.
3. **Note: This** optional bonus **homework will not affect your overall grade but offers extra** credit to improve your final score.
4. Due: **Monday Dec. 02, 2024 at 11:59pm**
5. No later submission is accepted.
6. Please read and follow submission instructions. No exception will be made to accommodate incorrectly submitted files/reports.
7. All students are required to typeset their reports using latex. Overleaf (`https://www.overleaf.com/learn/latex/Tutorials`) can be a good start.

---

1. (30 points) **Principal Component Analysis:**

   Report

   - Reconstruction error for p = 10 is 155351.4742

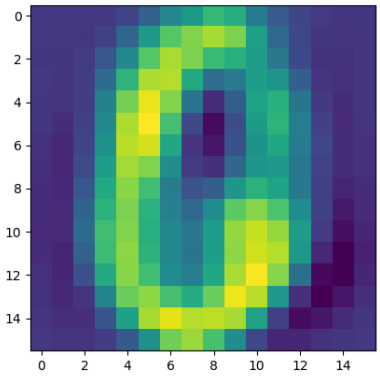

Figure 1: Image 1 with 10 component

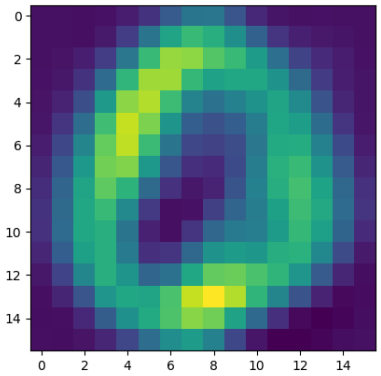

Figure 2: Image 2 with 10 component

   - Reconstruction error for p = 50 is 41024.8648

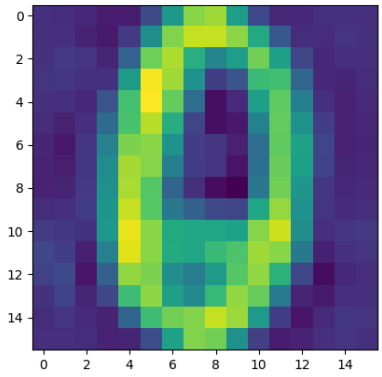

Figure 3: Image 1 with 50 component

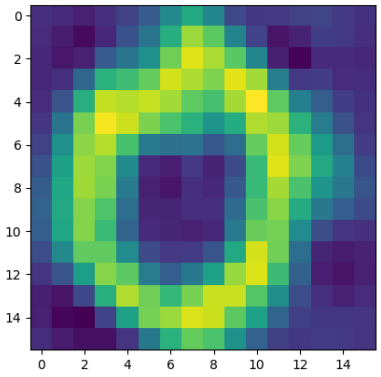

Figure 4: Image 2 with 50 component

- Reconstruction error for p = 100 is 14285.8125

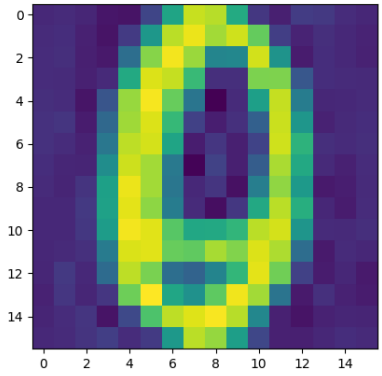

Figure 5: Image 1 with 100 component

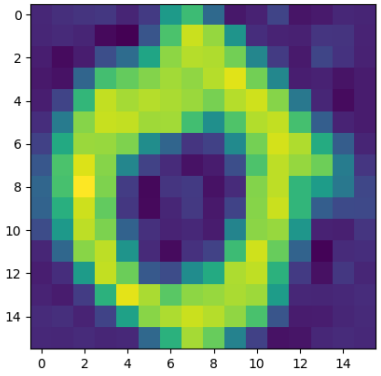

Figure 6: Image 2 with 100 component

- Reconstruction error for p = 200 is 1371.4145

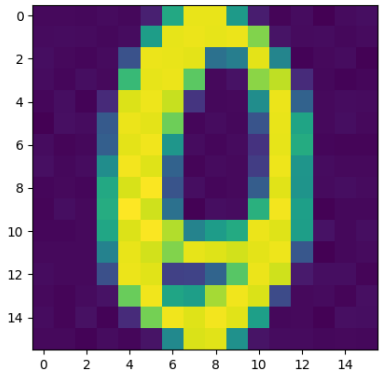

Figure 7: Image 1 with 200 component

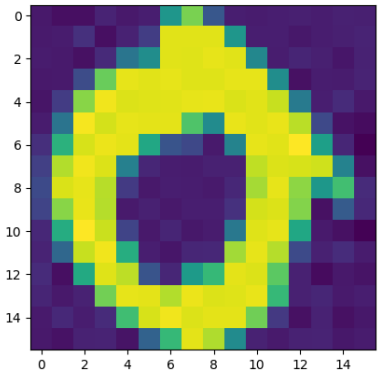

Figure 8: Image 2 with 200 component

As the number of principal components (p) increases, the reconstruction error decreases noticeably, showing that more components capture more variance from the original data. When $p = 10$, the error is high (155351.4742), and the reconstructed images lose a lot of detail. At $p = 50$, the error drops to 41024.8648, and the quality improves, but some finer details are still missing. For $p = 100$, the error is reduced further to 14285.8125, and the images look much closer to the originals. Finally, at $p = 200$, the error is very small (1371.4145), and the images are almost identical to the originals. This highlights a trade-off: higher p values give better quality but require more computational resources.