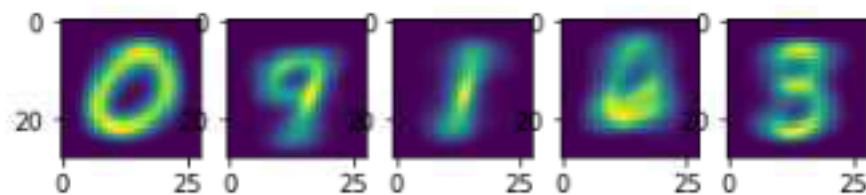


CS189/289A – Spring 2017 — Homework 7

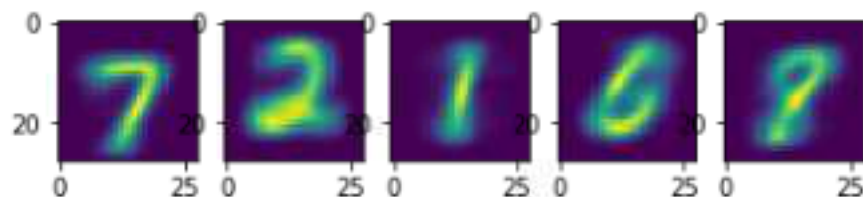
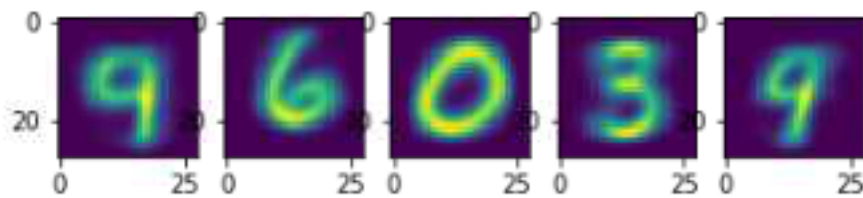
Yicheng Chen, SID 26943685

1. k-means clustering

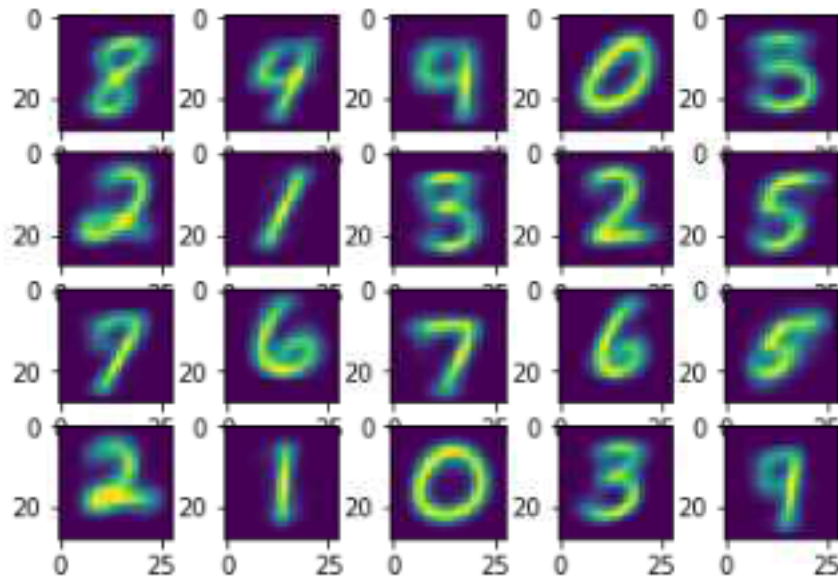
(a) The cluster centers of
k=5:



k=10:



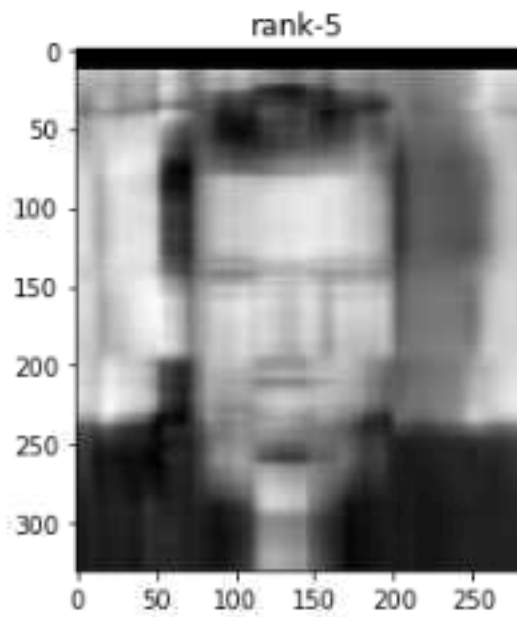
k=20:



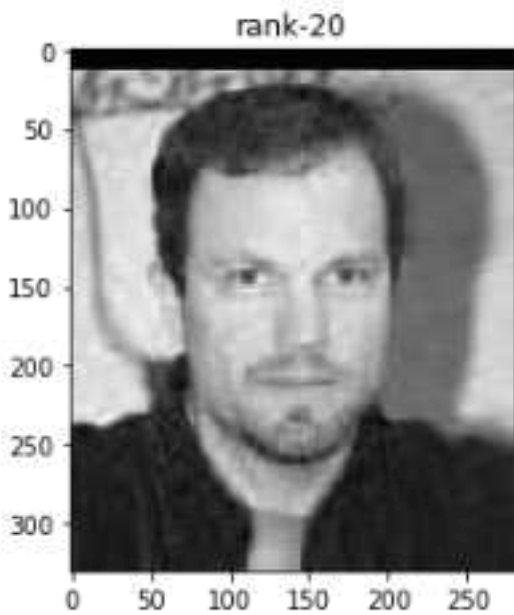
Where there are few centers, each center tends to represent more than 1 digits. And as the number of centers increases, different digits are better clustered, while due to the random initialization, there might be multiple centers representing the same digit.

2. Low rank approximation

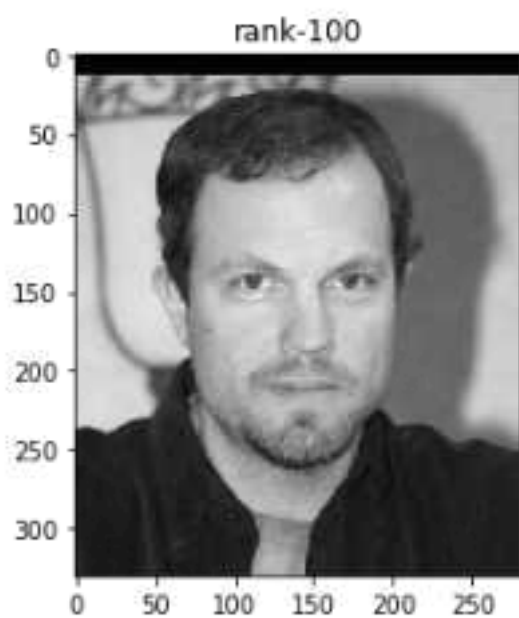
(a) Rank=5:



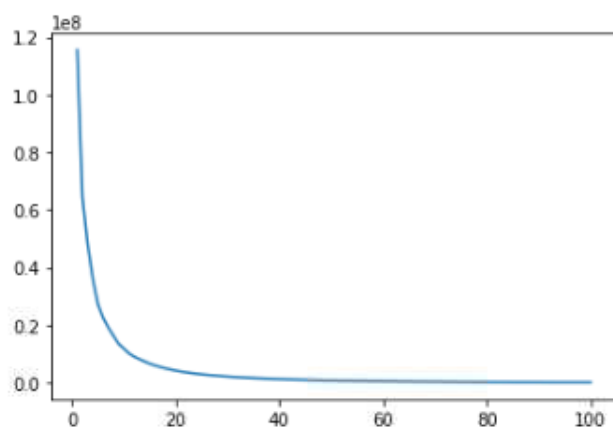
Rank=20:



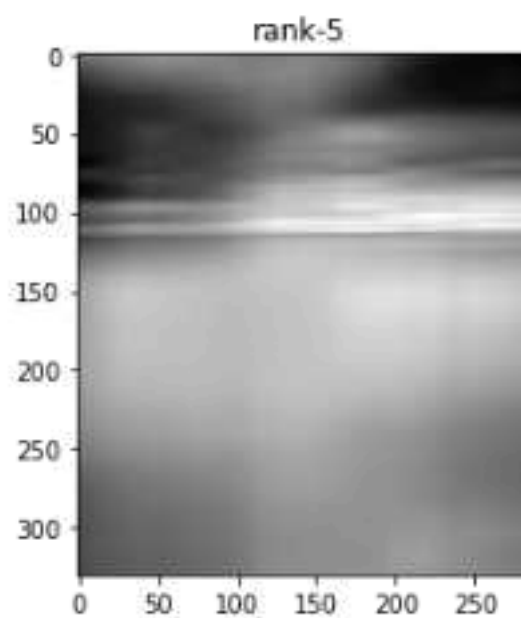
Rank=100:



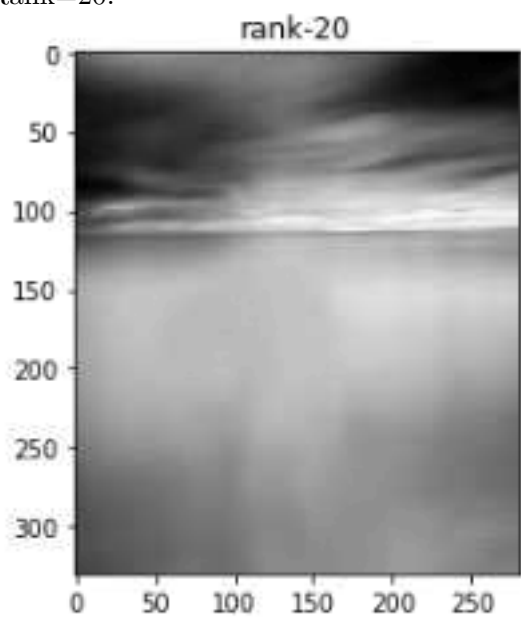
(b) The MSE:



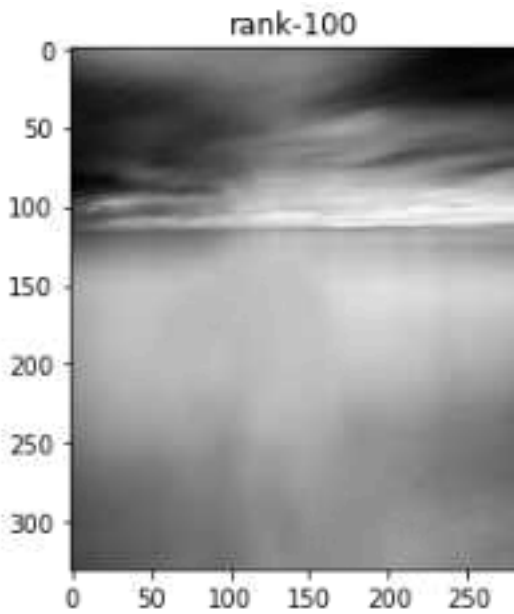
(c) Rank=5:



Rank=20:



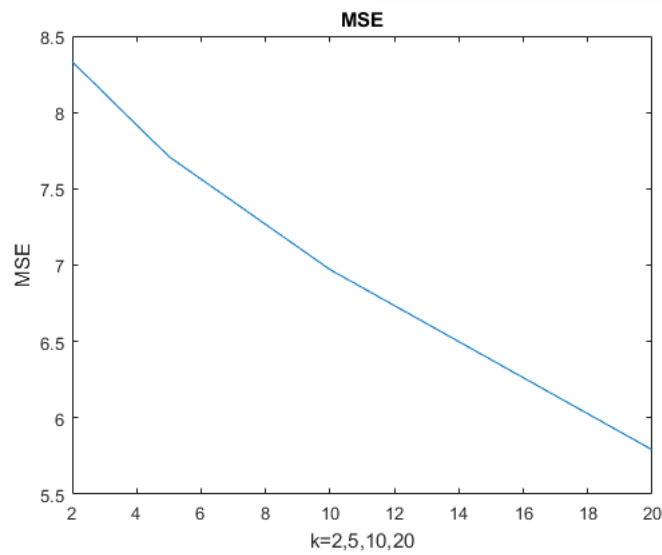
Rank=100:



- (d) For the face image, the lowest rank is about 50. However, for the sky image, the number is around 15. The sky image has significantly lower threshold because there are not as many details as in the face image, which requires more components to approximate.

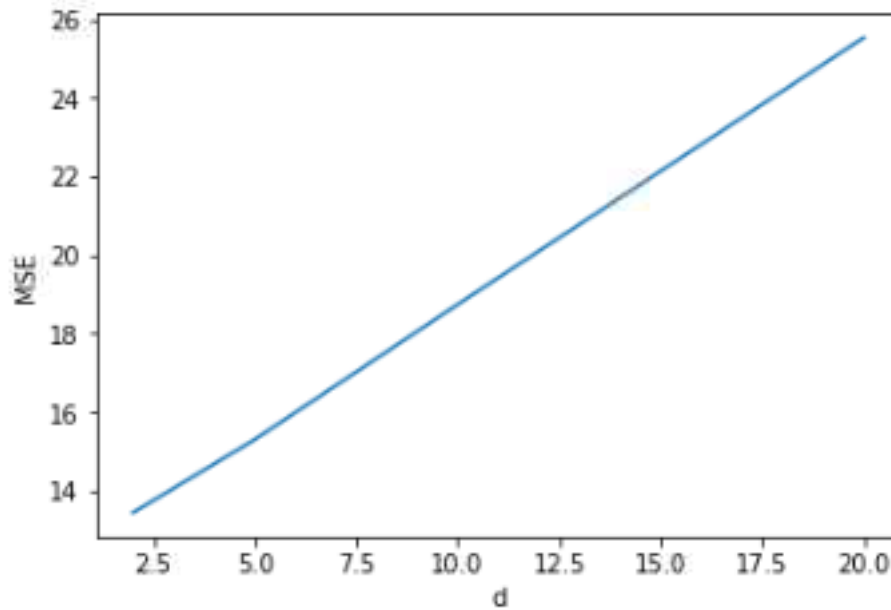
3. Joke Recommendation System

(a) MSE:



The validation accuracy is:
0.7051, 0.7154, 0.7165 and 0.6859,
respectively for $k=2, 5, 10$ and 20.

(b) MSE:



The validation accuracy is:
0.7054, 0.7103, 0.7203, 0.6797,
respectively for $k=2, 5, 10$ and 20.

(c) The Kaggle submission score: 0.72164
Kaggle display name: Yicheng Chen