CS2109S Tutorial 6

AY 25/26 Sem 1 — github/omgeta

A.

1. Since distortion is a non-increasing objective over a finite set of changes, it will eventually converge to 0.
2. K-Means++: Spread out initial starting points evenly
   PCA-Preprocess: Remove noise and give more accurate optimal
3. $\mu_1 = (1, 0.5)$, $\mu_2 = (2.5, 1.5)$
4. $\mu_1 = (1, 0)$, $\mu_2 = (2, 1)$
5. May need feature scaling, adjustment for non-convex shapes, non-linear shapes

B.

1. k weights for k dimension feature vectors
2. $k = 150$
3. Total reduced weights $= 10 \times 150 = 1500$
4. Original weights $= 10 \times 400 = 4000 \implies$ reduction factor of $\frac{150}{400} = 0.375$
5. It discards low importance features which are noise, decreases number of parameters early on for speed and storage
6. Natural images have strong correlations between neighbouring pixels. With random noise, the variance is uniformly spread across all directions. Therefore, there are no dominant principal components, so dropping features would reduce signal.