

דו"ח על ריצת האלגוריתם K-means

ביצעתי זאת באופן הבא:

- בחרתי באופן אקראי את ה-centroids, והרצתי את האלגוריתם k-means.
- בחרתי את פונקציית ה-Cost/avg loss הבאה:

$$Cost/avg = \frac{1}{m} \sum_{i=1}^m |z_i - x_i|^2$$

כלומר, חישוב של מרחק בריבוע בין כל פיקסל לבין ה-centroids הקרוב ביותר אליו, ולאחר מכן ממוצע על המרחקים. וחישבתי זאת בכל איטרציה.

ואלו התוצאות עבור כמות שונה של centroids:

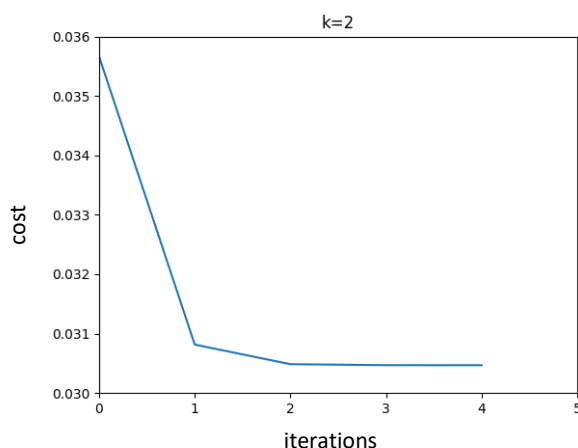
נשים לב כי בכל התוצאות, מאיטרציה אחת לשנייה יש ירידה (גם אם קטנה מאד) בערך ה-cost.

• 2 centroids

○ ה-centroids הם:

(0.809 0.4077 0.262), (0.7584 0.0523 0.0102)

○ גרף של פונקציית ה-cost/avg loss:



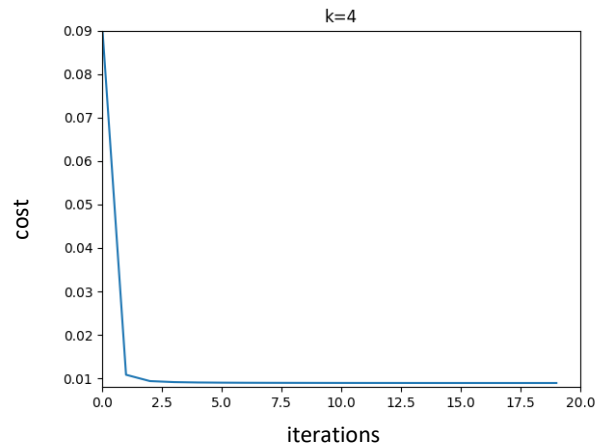
ניתן לשים לב שכאן יש התכנסות באיטרציה 4.

• 4 centroids

○ ה-centroids הם:

(0.8464 0.6856 0.5808), (0.8407 0.5508 0.4156),
(0.8508 0.4057 0.3333), (0.2636 0.3537 0.8583)

○ גרף של פונקציית ה- cost/avg loss:



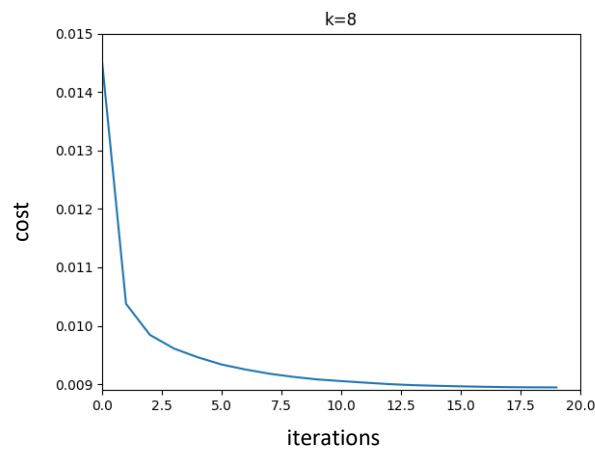
כאן אין התכנסות עד 20 איטרציות, ולכן האלגוריתם עצר ללא התכנסות.

• 8 centroids:

○ ה- centroids הם:

(0.7791 0.7704 0.0487), (0.4073 0.2175 0.6072), (0.3142 0.8488 0.9009),
(0.2968 0.6548 0.4602), (0.5975 0.2158 0.5456), (0.3723 0.7374 0.9246),
(0.1039 0.3972 0.1116), (0.565 0.2099 0.8278)

○ גרף של פונקציית ה- cost/avg loss:



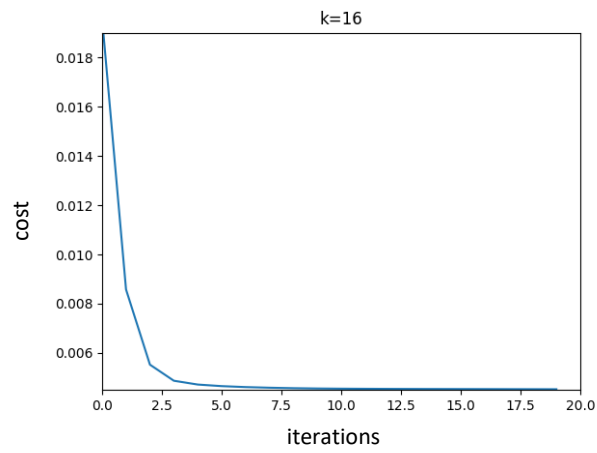
כאן אין התכנסות עד 20 איטרציות, ולכן האלגוריתם עצר ללא התכנסות.

• 16 centroids:

○ ה- centroids הם:

(0.8646 0.8272 0.1816), (0.1861 0.2101 0.5034), (0.1765 0.0383 0.5048),
(0.3054 0.3718 0.3021), (0.9009 0.1019 0.213), (0.5443 0.335 0.7052),
(0.6503 0.8034 0.4217), (0.6909 0.2309 0.9204), (0.088 0.8241 0.6927),
(0.8853 0.5744 0.2761), (0.702 0.9941 0.8735), (0.3841 0.1734 0.9295),
(0.4383 0.5147 0.087), (0.6309 0.1163 0.6268), (0.8608 0.0518 0.7468),
(0.6876 0.3673 0.9307)

○ גרף של פונקציית ה- cost/avg loss:



כאן אין התכנסות עד 20 איטרציות, ולכן האלגוריתם עצר ללא התכנסות.