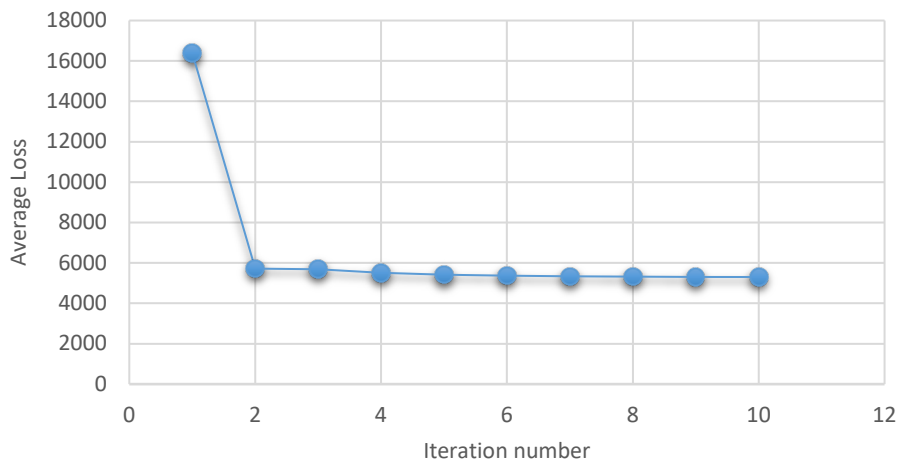


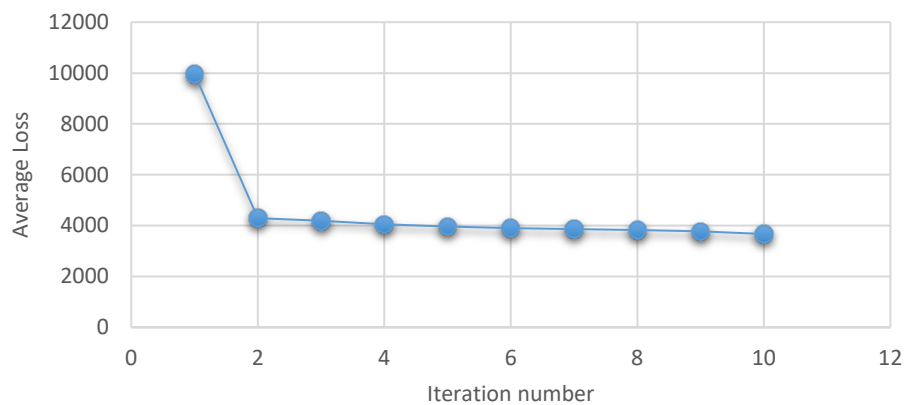
Report- Elad Israel 313448888

גרפי השגיאה הממוצעת
בהתאם למספר המרכזים:

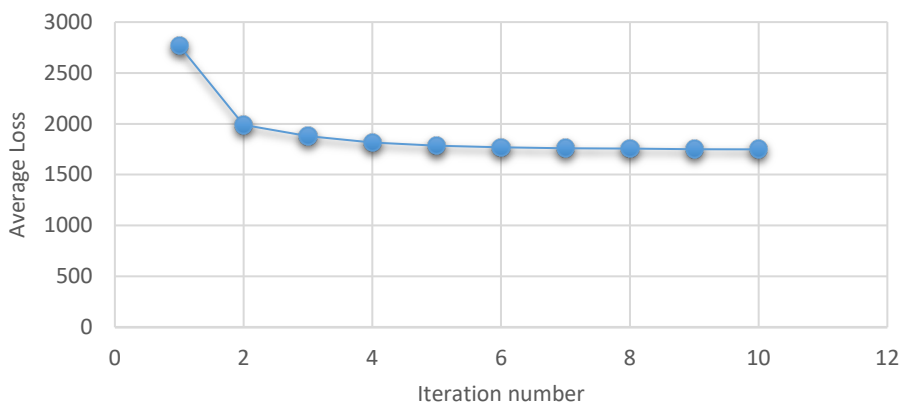
Average loss value as a function of the iterations with 2 centroids



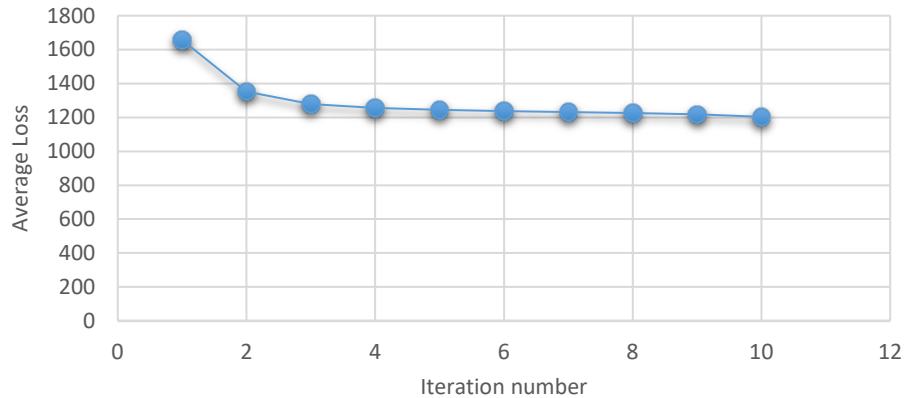
Average loss value as a function of the iterations with 4 centroids



Average loss value as a function of the iterations with 8 centroids



Average loss value as a function of the iterations with 16 centroids



הנקודות שבחרתי:

-15000 -15000 15000 15000	K=2
-15000 -15000 15000 15000 -10000 -10000 10000 10000	K=4
-15000 -15000 15000 15000 -10000 -10000 10000 10000 -5000 -5000 5000 5000 -2500 -2500 2500 2500	K=8
-15000 -15000 15000 15000 -10000 -10000 10000 10000 -5000 -5000 5000 5000 -2500 -2500 2500 2500 -500 -1500 1500 1500 -1000 -1000 1000 1000 -500 -500 500 500 -250 -250 250 250	K=16

הסבר מימוש:

בחרתי את המרכזים בהתאם לנקודות: ראיתי שהנקודות מתפרסות בסמוך לציר הישר $Y=X$, בטווחים של -30000 עד 30000 לערך. בהתאם לזה בחרתי את המרכזים שיתפרסו בצורה יחסית שווה לאורך הציר- כך שלכל מרכז יהיו נקודות, אבל לא יותר מידי נקודות למרכז ספציפי או פחות מידי.

הרצתי 10 איטרציות בהם עדכנתי את שיוך הנקודות למרכזים הקרובים ואז עדכנתי כל מרכז להיות ממוצע הנקודות המשויכות אליו.

בדקתי זאת עבור $K=2,4,8,16$ כאשר K זה כמות המרכזים, ואת תוצאות השגיאה ציינתי בגרפים לעיל.