

למידת מכונה - תרגיל בית 1

Report

רונלי ויגנסקי 211545892

התבקשנו לחשב את ממוצע ה loss כתלות באיטרציות הריצה. היינו צריכים לעשות זאת עבור $K=2,4,8,16$. כלומר עבור מספרים כאלו של צנטרואידים. לכן אתחלתי את הכמויות הללו של צנטרואידים בקבצים שונים בהתאמה לפי הקוד הבא:

```
import random

for i in range(1, 5):
    amount_cents = pow(2, i)
    file = open("cents" + str(amount_cents) + ".txt", "w")
    for j in range(amount_cents):
        file.write(
            str(round(random.uniform(0, 1), 4)) + " " + str(round(random.uniform(0, 1), 4)) + " " + str(
                round(random.uniform(0, 1), 4)) + "\n")
    file.close()
```

כפי שניתן לראות מהקוד, הצנטרואידים מאותחלים בצורה רנדומלית בין 0 ל 1 עם עיגול של 4 ספרות אחרי הנקודה.

כעת נרצה לחשב את ה loss בכל איטרציה. כלומר את סכום המרחקים המינימליים של כלל הפיקסלים מהצנטרואידים לפי הנוסחה בתרגול:

$$\sum_{i=1}^n \min_{j=1, \dots, k} ||x_i - z_j||_2^2$$

לבסוף חילקתי במספר הפיקסלים כדי לקבל את ה average loss כפי שהתבקשנו.

הוספתי קוד שעושה זאת, הקוד מצורף בתמונה הבאה:

```
def loss(pi, cents):
    s = 0
    for p in pi:
        min_dist = distance_for_loss(p, cents[0])
        for cent in cents:
            dis = distance_for_loss(p, cent)
            if dis < min_dist:
                min_dist = dis
        s = s + min_dist
    return s / len(pi)
```

השימוש בפונקציה זו מתבצע ע"י הקריאה הבאה:

```
for key in centroids.keys():
    if len(centroids[key]) != 0:
        # update the centroid
        z[key] = np.mean(centroids[key], axis=0).round(4)
        centroids[key] = []

s = loss(pixels, z)
```

כלומר לאחר עדכון הצנטרואיד "החדש" אנו מבצעים את החישוב של ה loss בכל איטרציה. כך, אנו מקבלים את ממוצע המרחקים של כל הפיקסלים הממופים לצנטרואיד המתאים להם בכל איטרציה.

נבנה גרף המציג את ממוצע ה loss כתלות באיטרציה. כתבתי גם פה קוד קצר שיבנה לי את הגרף בהתאם לנתונים על ה loss שחישבתי קודם. עבור כל k כתבתי כותרת מתאימה. לדוגמה נצרף עבור K = 2:

```
x = list(range(len(list_of_loss)))
plt.plot(x, list_of_loss, c='red')
plt.xticks(x, x)
plt.xlabel('Iteration')
plt.ylabel('Avg. loss')
plt.title('K = 2')
plt.show()
```

כך ש list_of_loss הוגדרה כרשימה ריקה שהוספתי אליה כל פעם את ערך ה loss שקיבלנו.

```
s = loss(pixels, z)
list_of_loss.append(s)
```

נשים לב שעבור K = 2 חלה התכנסות אחרי פחות מ-20 איטרציות.

ניתן לראות את מגמת הירידה בערכי ה avg loss כפי שצפינו שיקרה שכן עם הזמן מתקרבים להתכנסות, להלן הגרפים:



