APML | תרגיל 2

שם: אריאל ברוך | ת"ז: 205925704

25 בנובמבר 2020

1 Theoretical Part

1.1 MLE in the EM algorithm

התוחלת של הlog likelihood עבור Gaussian mixture model היא:

$$\mathbb{E}[l(S,\theta)] = \sum_{i=1}^{N}\sum_{y=1}^{k} c_{i,y} log(\pi_y N(x_i;\mu_y,\Sigma_y))$$

הראו כי הMLE של של המשקולות המעורבבים הוא אכן:

$$\pi_y = \frac{1}{N}\sum_{i=1}^{N} c_{i,y}$$

פתרון:

כדי להוכיח את הטענה נרצה להשתמש בשיטת כופלי לאגראנז'. נרצה לעמוד בשני התנאים הבאים:

$$(1)\ \pi = argmax_{\pi\in\mathbb{R}}\left\{\sum_{i=1}^{N}\sum_{y=1}^{k} c_{i,y} log(\pi_y N(x_i;\mu_y,\Sigma_y))\right\} = argmax_{\pi\in\mathbb{R}}\left\{\sum_{i=1}^{N}\sum_{y=1}^{k} c_{i,y} log(\pi_y) + c_{i,y} log(N(x_i;\mu_y,\Sigma_y))\right\}$$

$$(2)\ \sum_{y=1}^{k}\pi_y = 1$$

ראשית, נבחין כי בביטוי (1) נוכל להתעלם מהביטוי הימני כי הוא לא פונקציה של π. כעת:

$$L(x,y,\lambda) = \sum_{i=1}^{N}\sum_{y=1}^{k} c_{i,y} log(\pi_y) + \lambda\left(1-\sum_{y=1}^{k}\pi_y\right)$$

$$\frac{\partial L}{\partial \pi_y} = \frac{\sum_{i=1}^{N} c_{i,y}}{\pi_y} - \lambda \quad\Rightarrow\quad \lambda = \frac{\sum_{i=1}^{N} c_{i,y}}{\pi_y} \quad \forall y\in[k]$$

כדי שמה שקיבלנו יהיה נכון, על λלהיות בפרופורציה ל $\sum_{i=1}^{N} c_{i,y}$ ולכן:

$$\pi_y = \frac{\sum_{i=1}^{N} c_{i,y}}{\sum_{y=1}^{k}\sum_{i=1}^{N} c_{i,y}} = \frac{\sum_{i=1}^{N} c_{i,y}}{N}$$

כאשר המעבר האחרון נובע מכך שמההגדרה: $\sum_{y=1}^{k} c_{i,y} = 1$ ואז $\lambda = N$

1.2 MLE in GSM algorithm

מודל GSM מניח שמקטעי התמונות הן דגימות מתוך k גאוסיאנים עם התפלגות $N(0, r_y^2\Sigma)$ עבור Σ קבועה בין הגאוסיאנים.
הראו כי העדכון המקסימלי עבור r_y^2 הוא:

$$r_y^2 = \frac{\sum_{i=1}^{n} c_{i,y} x_i^T \Sigma^{-1} x_i}{d\sum_{i=1}^{n} c_{i,y}}$$

פתרון:
בעיית האופטימיזציה היא

$$r_y = argmax_{r_y \in \mathbb{R}} \left\{ \sum_{i=1}^{N}\sum_{y=1}^{k} c_{i,y} log(\pi_y N(x_i; 0, r_y^2 \mathit{\Sigma}_y)) \right\} = argmax_{r_y \in \mathbb{R}} \left\{ \sum_{i=1}^{N}\sum_{y=1}^{k} c_{i,y} log(\pi_y) + c_{i,y} log(N(x_i; 0, r_y^2 \mathit{\Sigma}_y)) \right\}$$

נוכל גם כאן להוריד את המחובר השמאלי שלא כולל את r_y. נקבל:

$$r_y = argmax_{r_y \in \mathbb{R}} \left\{ \sum_{i=1}^{N}\sum_{y=1}^{k} c_{i,y} log(N(x_i; 0, r_y^2 \mathit{\Sigma}_y)) \right\}$$

נזכור כי:

$$N(x_i; 0, r_y^2 \mathit{\Sigma}_y) = \frac{1}{\sqrt{|2\pi r_y^2 \Sigma_y|}} e^{-\frac{1}{2}(x_i)^T r_y^2 \Sigma_y^{-1}(x_i)}$$

נפעיל log ונקבל:

$$logN(x_i; 0, r_y^2 \mathit{\Sigma}_y) = -\frac{1}{2} log(det(2\pi r_y^2 \Sigma_y)) - \frac{1}{2}(x_i)^T r_y^{-2} \Sigma_y^{-1}(x_i) \underset{*}{=}$$

$$-\frac{2d}{2} log(r_y) + log(det(2\pi \Sigma_y)) - \frac{1}{2}(x_i)^T r_y^{-2} \Sigma_y^{-1}(x_i)$$

* כי $det(cA) = c^d det(A)$ כאשר A מטריצה ממימד $d \times d$
נשים לב שהביטוי האמצעי לא כולל את r_y, לכן נרצה למקסם את הביטוי:

$$r_y = argmax_{r_y \in \mathbb{R}} \left\{ \sum_{i=1}^{N}\sum_{y=1}^{k} c_{i,y} \left(-d log(r_y) - \frac{1}{2r_y^2}(x_i)^T \Sigma_y^{-1}(x_i) \right) \right\}$$

נגזור ונשווה לאפס. נקבל לכל $y \in [k]$:

$$\sum_{i=1}^{N} c_{i,y}\left(-\frac{d}{r_y} + \frac{1}{r_y^3}(x_i)^T\Sigma_y^{-1}(x_i)\right) = 0$$

$$\sum_{i=1}^{N} c_{i,y}\left((x_i)^T\Sigma_y^{-1}(x_i)\right) = \frac{r_y^3\sum_{i=1}^{N} c_{i,y}d}{r_y}$$

$$r_y^2 = \frac{\sum_{i=1}^{n} c_{i,y}x_i^T\Sigma^{-1}x_i}{d\sum_{i=1}^{n} c_{i,y}}$$

כנדרש.

1.3 EM Initialization

יוסי אתחל את EM כך:

$$\forall y:\ \pi_y = \frac{1}{k},\ \mu_y = \mu,\ \Sigma_y = \Sigma$$

כתבו את שתי האיטרציות הראשונות.

פתרון:

איטרציה 1:

חישוב ה log likelihood:

$$c_{i,y} = \frac{\pi_y N(x_i;\mu_y,\Sigma_y)}{\sum_{l=1}^{k}\pi_l N(x_i;\mu_l,\Sigma_l)} = \frac{\frac{1}{k}N(x_i;\mu_y,\Sigma_y)}{k\frac{1}{k}N(x_i;\mu_y,\Sigma_y)} = \frac{1}{k}$$

עדכון הפרמטרים:

$$\pi_y = \frac{1}{N}\sum_{i=1}^{N} c_{i,y} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{k} = \frac{1}{k}$$

$$\mu_y = \frac{\sum_{i=1}^{N} c_{i,y}x_i}{\sum_{i=1}^{N} c_{i,y}} = \frac{\sum_{i=1}^{N}\frac{1}{k}x_i}{\sum_{i=1}^{N}\frac{1}{k}} = \frac{\sum_{i=1}^{N} x_i}{N} = \bar{x}$$

$$\Sigma_y = \frac{\sum_{i=1}^{N} c_{i,y}(x_i-\mu_y)(x_i-\mu_y)^T}{\sum_{i=1}^{N} c_{i,y}x_i} = \frac{\sum_{i=1}^{N}\frac{1}{k}(x_i-\bar{x})(x_i-\bar{x})^T}{\frac{N}{k}} = \frac{\sum_{i=1}^{N}\frac{1}{k}(x_i-\bar{x})(x_i-\bar{x})^T}{N} = \overline{\Sigma}$$

איטרציה 2:

חישוב ה log likelihood:

$$c_{i,y} = \frac{\pi_y N(x_i;\mu_y,\Sigma_y)}{\sum_{l=1}^{k}\pi_l N(x_i;\mu_l,\Sigma_l)} = \frac{1}{k}$$

עדכון הפרמטרים:

$$\pi_y = \frac{1}{N}\sum_{i=1}^{N} c_{i,y} = \frac{1}{k}$$

$$\mu_y = \frac{\sum_{i=1}^{N} c_{i,y} x_i}{\sum_{i=1}^{N} c_{i,y}} = \bar{x}$$

$$\Sigma_y = \frac{\sum_{i=1}^{N} c_{i,y}(x_i-\mu_y)(x_i-\mu_y)^T}{\sum_{i=1}^{N} c_{i,y} x_i} = \overline{\Sigma}$$

כלומר, הפרמטרים נשארו כפי שהם ואין עדכון, לכן נרצה שיהיה עירבוב של גאוסיאנים כדי לשפר את יכולת ההבעה של האלגוריתם.

2 Practical Part

אופן ההרצה: עם ערך $k=5$ ועבור $epsilon = 0.01, 0.05, 0.1, 0.2$ קבועי רעש. גודל כל patch הוא $[8,8]$ מתוך התמונה.

2.1 זמני ריצה:

זמן הריצה של המודל GSM משמעותית יותר ארוך מMVN כפי שניתן לראות בטבלה הבאה:

	MVN	GSM
learning[sec]	0.0069	4.55
denoising[sec]	89.90	812.06

2.2 :log likelihood

	MVN	GSM
log likelihood	$1,428,508.688$	$2,012,762.629$

2.3 :MSE

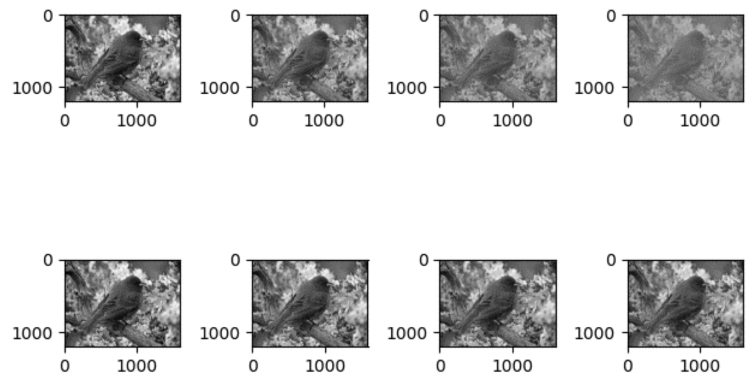
ביצועי המודל MVN טובים במעט יותר מאשר GSM:

epsilon	MVN	GSM
0.01	$7.258e-05$	0.0021
0.05	0.00054	0.0443
0.1	0.0010	0.0419
0.2	0.0018	0.0522

2.4 :denoised images

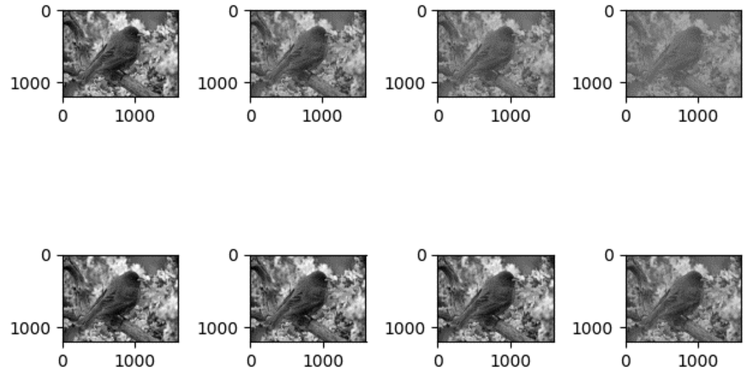
2.4.1 MVN

אפשר להבחין כי MVN מצליח להתמודד עם הרעש בכל התמונות:



2.4.2 GSM

אפשר להבחין כי הGSM מצליח להוריד את הרעש ברוב התמונות למעט התמונה הרועשת ביותר (ימני ביותר) בה יש הבדל נראה לעין לעומת שאר התמונות:



2.5 סיכום:

נבחין כי במרבית הפרמטרים שנבדקו (זמן ריצה, MSE ואיכות שיפור התמונות) אלגוריתם MVN מחזיר תוצאות יותר טובות מGSM. ההבדלים באיכות התוצאה לא מאוד משמעותיים, אבל עצם העובדה שמדובר במודל פשוט יותר שמביא תוצאות טובות יותר הופכת אותו למודל עדיף בהרבה לשימוש. להבנתי מהשוואה מול חברי הקבוצה יש אנשים שקיבלו תוצאות יותר טובות בMVN ואחרים בGSM. הייתי מצפה לקבל תוצאות טובות בGSM שכן הוא מודל מורכב יותר שכולל מספר גאוסיינים, ולכן יוכל תיאורטית ללמוד את התפלגות הרעש בצורה טובה יותר ואילו MVN מודל פשוט של גאוסיאן אחד. יתכן שההבדלים בביצועי המודלים תלויי מימוש.