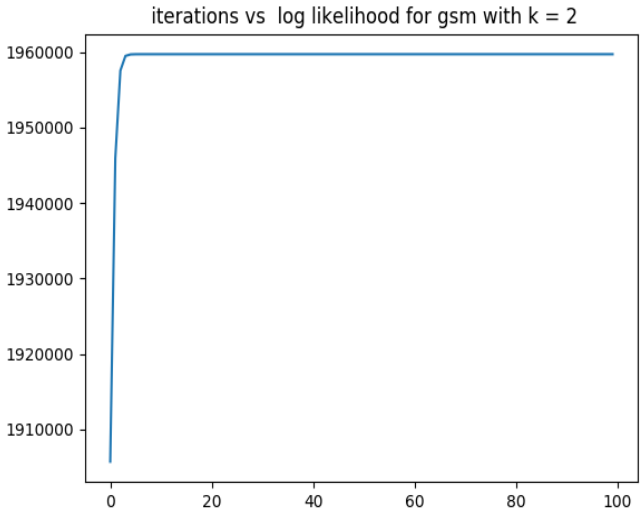
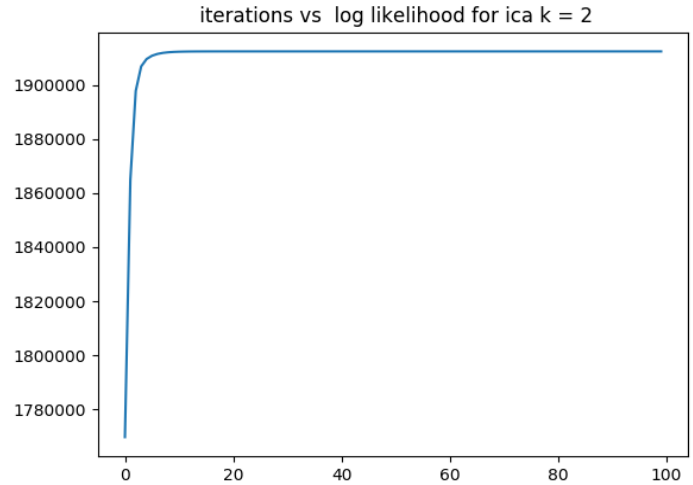
Image Denoising
id : 305248791
user : chris91

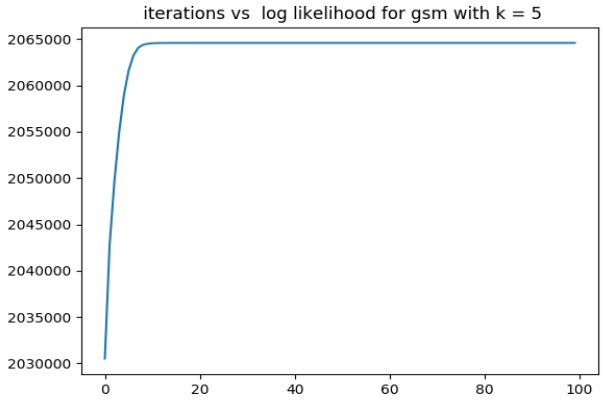
Note: the code takes time to run because it tested on all test images, and averaging the MSE on the whole set. And the code creates several plots.
If the code is run and tested on one image it won't take long time at all.

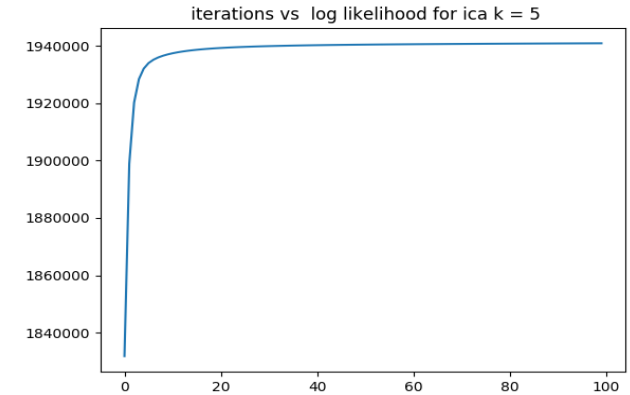
2. Write a function for learning each model using the EM algorithm. Show plots of the log likelihood of the data as a function of the EM iterations and show EM indeed converges.

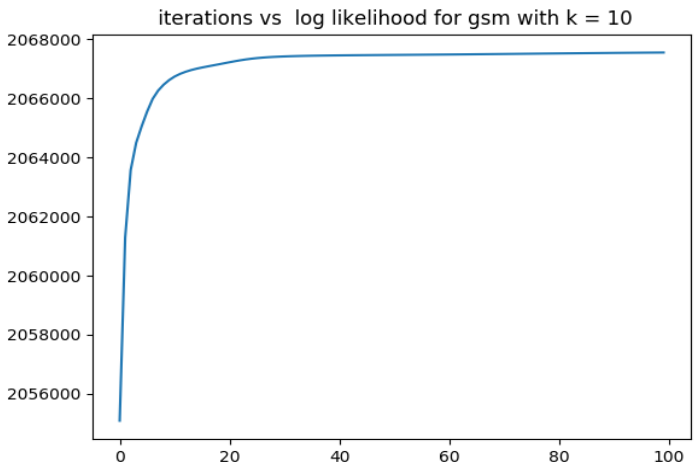
log likelihood on train data for mvn = 1443162.1740455246

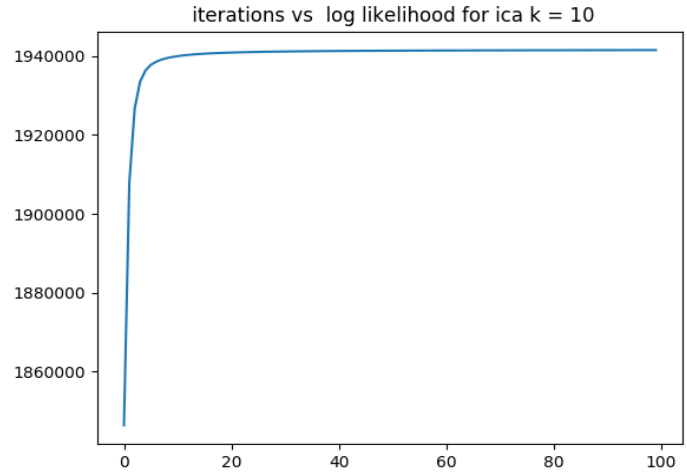






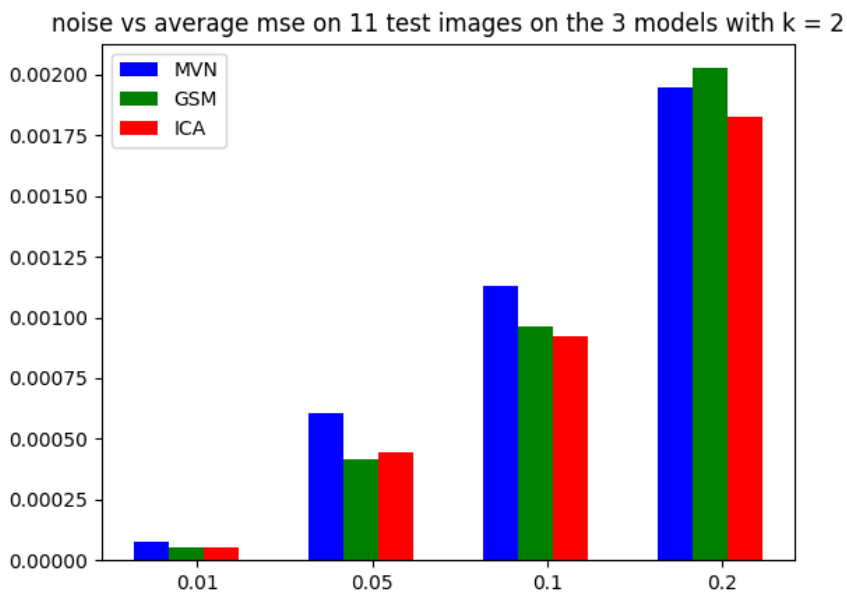


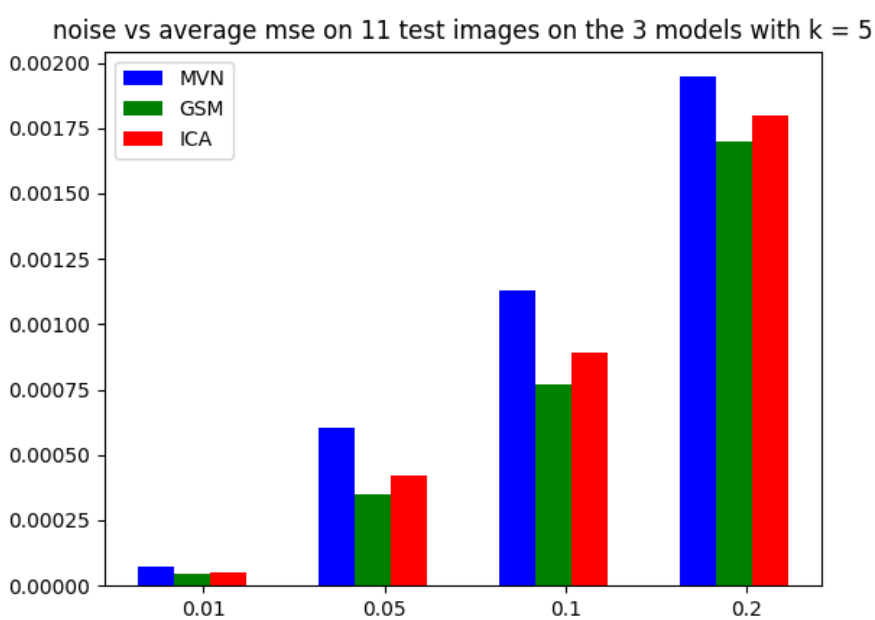


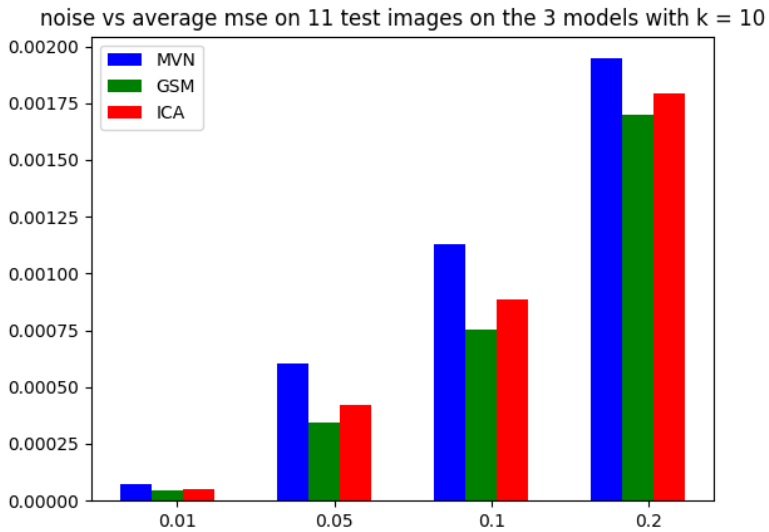


4. Provide a comparison of the three models in terms of run time, the log-likelihood of the clean test set and the MSE of the reconstruction of the test set images. Also, write which model you think is best and why. Provide qualitative evidence to support your conclusions (i.e. noisy and denoised images).

I averaged the MSE on 11 test images for noise range = (0.01, 0.05, 0.1, 0.2) , iterations on ICA and GSM model is 100, the k's (The number of components) are k = [2, 5, 10].





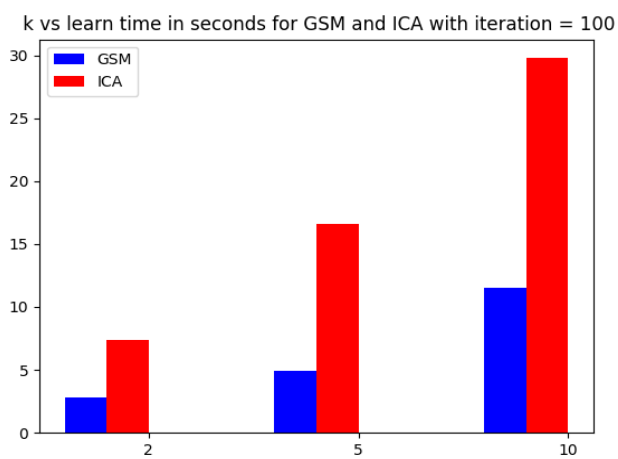


with k = 2 ICA is better than GSM only on noise 0.1 and 0.2.
for k = 5 and k = 10 GSM is the best model, while the ICA comes second.
lets look at the values for ICA with k=2, where it performs better on some of the noises than GSM and MVN, and compare those values with GSM on k = 5 and k =10 where GSM performs better than MVN and ICA for all noises.

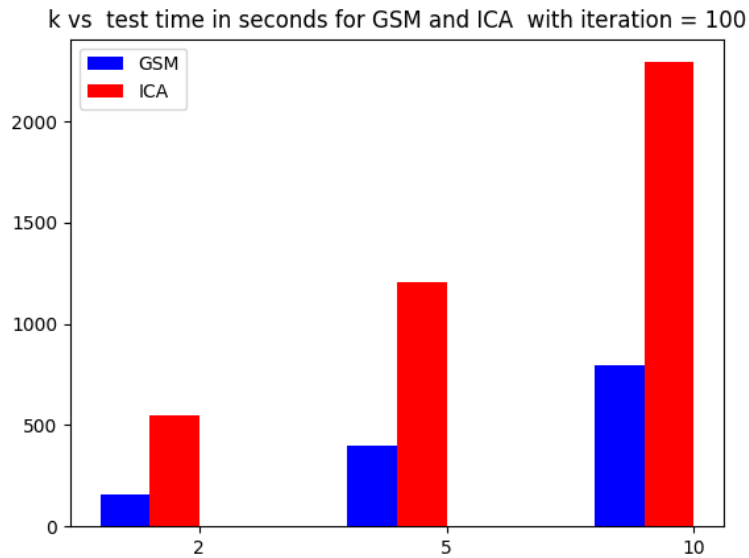
Noise	0.01	0.05	0.1	0.2
Average MSE ICA k = 2	5.35401973e-05	4.38360388e-04	9.11455635e-04	1.80156318e-03
Average MSE GSM k =5	4.59689014e-05	3.51971399e-04	7.67302624e-04	1.70443444e-03
Average MSE GSM K = 10	**4.52713133e-05**	**3.44413159e-04**	**7.56453514e-04**	**1.69813761e-03**

it is clear now that GSM with k = 5 and k = 10 performs better than ICA with all noise values.
conclusion GSM with k = 10 is the best model.

Now we look at learning time:
learn time for mvn = 0.004167079925537109.
as for GSM and ICA:

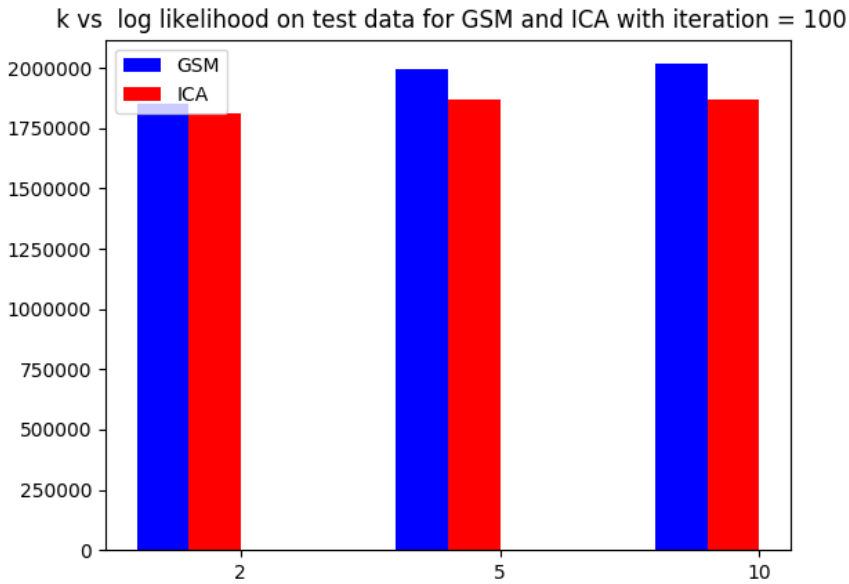


naturally mvn is the fastet in learn time, GSM is faster than ICA .
As for test time, I calculated the sum of time on denoising 11 images on four different noises, for ICA and GSM it was done, on different k values.
overall test time for MVN on 11 test images with four different noises on each image = 26.414379596710205 seconds



naturally mvn is fastet on test time , gsm second and ica is the slowest.

As for the loglikelihood on test data:
log likelihood on test data for mvn = 1420208.3094142738
log likelihood on test data for ICA and GSM for different k's:

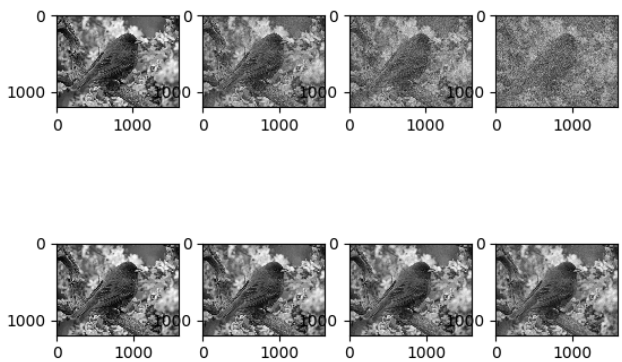


closer look at log likelihood on test data:

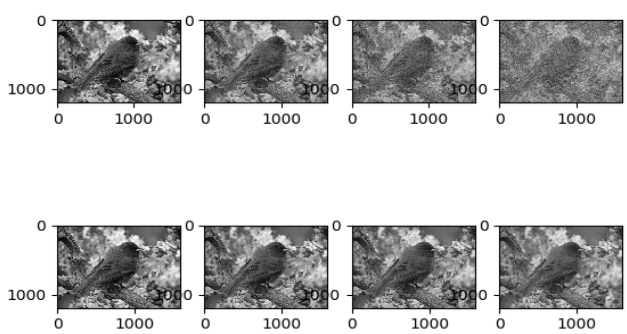
k	2	5	10
log likelihood on test data for GSM	1852567.0577450194	1997923.234907756	2016984.5669015718
log likelihood on test data for ICA	1813554.5359805862	1866935.7870219892	1867710.6429295682

both ICA and GSM get better results than MVN while, GSM performs better on each k, with **GSM with k = 10 is shown to be the best model, this is consistent with average mse values also there GSM with k = 10 performed best.**

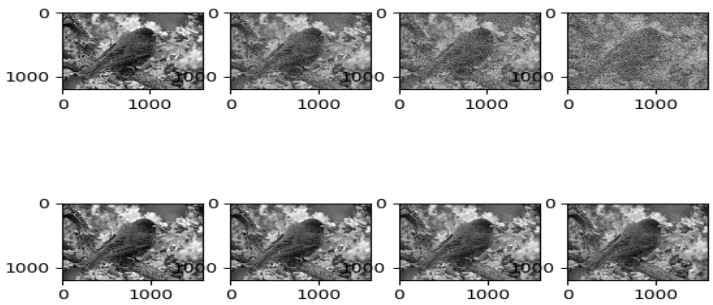
Examples of desnoised images.
Mvn :



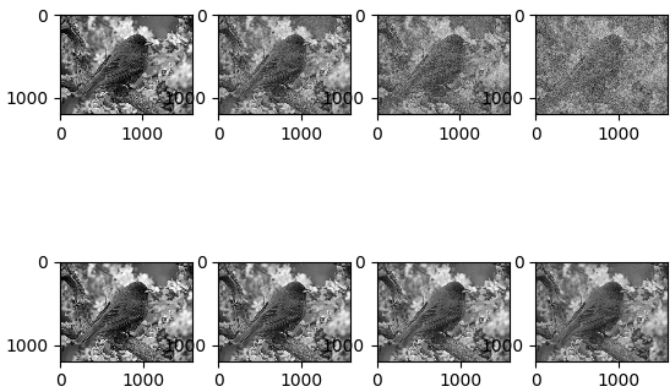
gsm with k = 2 , learned with iteration = 100:



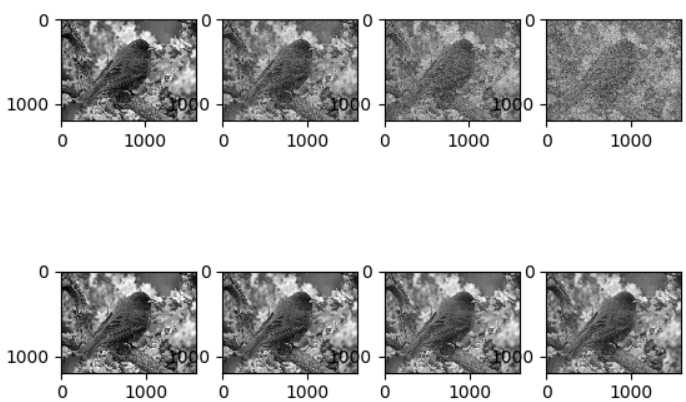
ica with k = 2 , learned with iteration = 100:



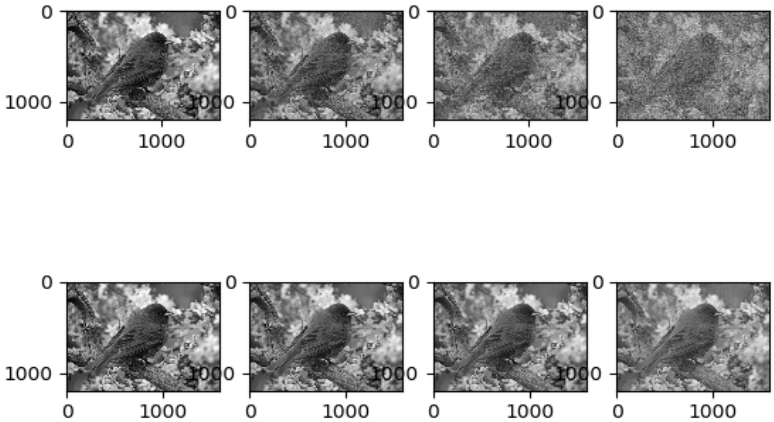
gsm with k = 5 , learned with iteration = 100:



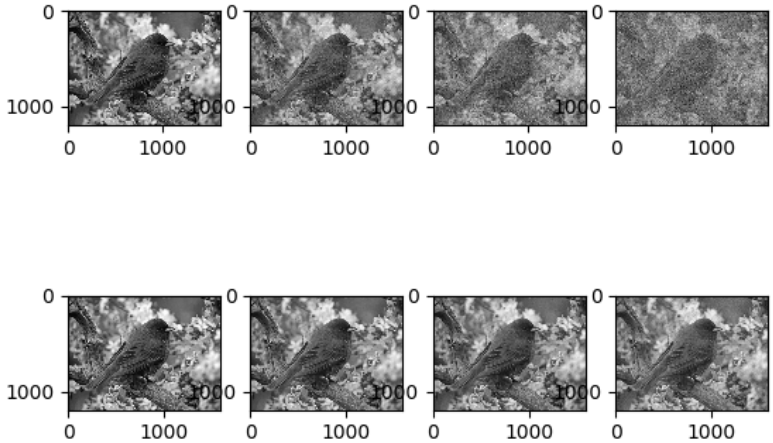
ica with k = 5 , learned with iteration = 100:



gsm with k = 10 , learned with iteration = 100: best model



ica with k = 10 , learned with iteration = 100:



changed functions:
test_denoising_with_plots
it plots the the images vs denoised images, used to show example of
denoising on image.

test_denoising , I turned off the printing and plotting and it now returns:
:return:sum_denoise_time, overall denoise time on all noise range .
:return: mse_denoised_array the mse of the denoised image for each noise.

learn_GSM the return is changed
:return: A list of trained GSM_Model objects, a model for each iteration
the last element would be the best model .

learn_ICA now returns
:return: A trained ICA_Model object.
:return: all_mix, all mix values for each iteartion. shape = (iter_num, d,k)
:return: all_vars, all vars values for each iteration.shape = (iter_num, d,k)

שאלות תיאורטיות.

שאלה 1.1:

לכל y π_y מתפלג בהתפלגות מולטינומית
ומתקיימים הקשרים הבאים:

$$\pi_y \geq 0$$

$$\sum_{y=1}^{k} \pi_y = 1$$

נגדיר את הלגראנז'ן:

$$L(\zeta,\theta,\lambda) = \sum_{i=1}^{n}\sum_{y=1}^{k} c_{iy}\left(\log(\pi_y) + \log(N(x_i,\mu_y,\Sigma_y))\right) - \lambda(\pi^T 1 - 1)$$

נגזור לפי π_y:

$$\frac{\partial L}{\partial \pi_y} = \sum_{i=1}^{n}\frac{c_{iy}}{\pi_y} - \lambda = 0 \Rightarrow \pi_y = \frac{\sum_{i=1}^{n} c_{iy}}{\lambda}$$

נשתמש בתנאי $\sum_{i=1}^{k}\pi_y = 1$ כדי לקבל את λ:

$$\sum_{i=1}^{k}\pi_y = 1 \Rightarrow \underbrace{\sum_{i=1}^{k}\sum_{i=1}^{n} c_{iy}}_{\lambda} = 1 \Rightarrow \lambda = \cancel{\sum_{i=1}^{k}\sum_{i=1}^{n} c_{iy}}$$

$$\cancel{\pi_y = \sum_{i=1}^{n} c_{iy}}$$

$$\Downarrow$$

$$\lambda = \sum_{i=1}^{k}\sum_{i=1}^{n} c_{iy}$$

נציב ונקבל:

$$\pi_y = \frac{\sum_{i=1}^{n} c_{iy}}{\underbrace{\sum_{i=1}^{k}\sum_{y=1}^{k} c_{iy}}_{=1}} = \frac{1}{n}\sum_{i=1}^{n} c_{i,y}$$

Scanned by CamScanner

1.3 סעיף ב:

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

$$N(x, \vec{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

1.3 ג' ב'ל, EM מ/ל/ה

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

$$= \frac{\frac{1}{k} \cdot \frac{e^{-\frac{1}{2}(x_i-\mu_y)\Sigma_y^{-1}(x_i-\mu_y)}}{\sqrt{(2\pi)^d |\Sigma_y|}}}{\sum_{l=1}^{k} \frac{1}{k} \cdot \frac{e^{-\frac{1}{2}(x_i-\mu_l)\Sigma_l^{-1}(x_i-\mu_l)}}{\sqrt{(2\pi)^d |\Sigma_y|}}}$$

$$= \frac{\cancel{\frac{1}{k}} \cdot \frac{e^{-\frac{1}{2}(x_i-\mu)\Sigma^{-1}(x_i-\mu)}}{\cancel{\sqrt{(2\pi)^d |\Sigma|}}}}{\cancel{\frac{1}{k}} \sum_{l=1}^{k} \frac{e^{-\frac{1}{2}(x_i-\mu)\Sigma^{-1}(x_i-\mu)}}{\cancel{\sqrt{(2\pi)^d |\Sigma|}}}} = \frac{1}{k} = C$$

לא תלוי ב x_i

לא תלוי ב l

$$\pi_y = \frac{1}{N}\sum_{i=1}^{N} c_{iy} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{K} = \frac{1}{K} = \pi$$

ביטוי זהה לסמן ב-π.

$$\mu_y = \frac{\sum_{i=1}^{N} c_{iy} x_i}{\sum_{i=1}^{N} c_{iy}} = \frac{\sum_{i=1}^{N}\frac{1}{K} x_i}{\sum_{i=1}^{N}\frac{1}{K}} = \frac{\frac{1}{K}\sum_{i=1}^{N} x_i}{\frac{1}{K}\cdot N}$$

$$= \frac{\sum x_i}{N} = \hat{\mu}$$ (ממוצע בכל המדגם, ביטוי זהה) תוחלת

$$\Sigma_y = \frac{\sum_{i=1}^{N} c_{iy}(x_i-\mu_y)(x_i-\mu_y)^T}{\sum_{i=1}^{N} c_{iy}} = \frac{\sum_{i=1}^{N}\frac{1}{K}(x_i-\hat{\mu})(x_i-\hat{\mu})^T}{\sum_{i=1}^{N}\frac{1}{K}}$$

$$= \frac{\frac{1}{K}\sum_{i=1}^{N}(x_i-\hat{\mu})(x_i-\hat{\mu})^T}{\frac{1}{K}\cdot N} = \frac{\sum_{i=1}^{N}(x_i-\hat{\mu})(x_i-\hat{\mu})^T}{N} = \hat{\Sigma}$$

אותו בכל המדגם ביטוי זהה

נציב בנוסחה:

$$c_{iy} = \frac{\pi \cdot \frac{1}{\sqrt{(2\pi)^d\,|\Sigma|}} \cdot e^{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1} (x_i-\mu)}}{\sum_{i=1}^{k} \pi \frac{e^{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1} (x_i-\mu)}}{\sqrt{(2\pi)^d\,|\Sigma|}}}$$

$$= \frac{1}{k} = c$$

$$\pi_y = \frac{1}{N}\sum_{i=1}^{N} c_{iy} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{k} = \frac{1}{k} = \pi$$

$$\mu_y = \frac{\sum_{i=1}^{N} c_{iy} x_i}{\sum_{i=1}^{N} c_{iy}} = \frac{\sum_{i=1}^{N}\frac{1}{k} x_i}{\sum_{i=1}^{N}\frac{1}{k}} = \frac{\sum_{i=1}^{N} x_i}{N} = \hat{\mu}$$

$$\Sigma_y = \frac{\sum_{i=1}^{N} c_{iy}(x_i-\mu_y)(x_i-\mu_y)^T}{\sum_{i=1}^{N} c_{iy}} = \frac{\sum_{i=1}^{N}\frac{1}{k}(x_i-\hat{\mu})(x_i-\hat{\mu})^T}{\sum_{i=1}^{N}\frac{1}{k}}$$

$$= \frac{\sum_{i=1}^{N}(x_i-\hat{\mu})(x_i-\hat{\mu})^T}{N} = \hat{\Sigma}$$

~~[illegible]~~

נבדוק שהם גאוסיאנים הסתברויות שווים לקטגוריה

לכן כדי לדגום אז יש תרגילים. נצטרך הגרלה בין

הגרלות של תוחלת וסטיות תקן לגאוסיאנים:

$$P(x) = \sum_{y=1}^{k} \pi_y N(x;\mu_y,\Sigma_y) = \sum_{y=1}^{k}\frac{1}{k} N(x,\hat{\mu},\hat{\Sigma}) = N(x,\hat{\mu},\hat{\Sigma})$$

Scanned by CamScanner

לכן, סקלר לורנץ בהגבלות של באוסין

אחד עם תוחלת בהגבלה $\hat{u}=u_y$,

וקוואריאנס בהגבלה $\xi=y_\xi$. ~~[illegible]~~ ~~[illegible]~~

~~[illegible]~~ ולכן לא שקול למודל

עם ~~[illegible]~~ באוסין אחד ויחיד

ולכן מוכן של אפר ובתיי באוסינים

ובאות בנפרד.

Scanned by CamScanner

שאלה 1.2:

נרחיב את ה loglikelihood של הערבוב
באופן הבא:

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

בנוסף נניח ש: $\Sigma_y = r_y^2 \Sigma$ עבור Σ משותף
כך הנורמליזציה ! $\mu_y = 0$.

לכן שנשתמש כדי לבטא את הפרמטרים r_y באופן מינימלי
נכפיל את הביטוי:

$$E[l(S,\theta)] = \sum_{i=1}^{N}\sum_{y=1}^{K} c_{i,y}\log(\pi_y) + c_{i,y}\log N(x_i; 0, r_y^2\Sigma)$$

נחליף את סדר הסכימות:

$$= \sum_{y=1}^{K}\left(\sum_{i=1}^{N} c_{i,y}\log(\pi_y) + \sum_{i=1}^{N} c_{i,y}\log\left(\frac{1}{\sqrt{|2\pi r_y^2\Sigma|}}\cdot e^{-\frac{1}{2}x_i^T(r_y^2\Sigma)^{-1}x_i}\right)\right)$$

$$= \sum_{y=1}^{K}\left(\sum_{i=1}^{N} c_{i,y}\log(\pi_y) + \sum_{i=1}^{N} c_{i,y}\log\left(e^{(-1/2r_y^2)\, x_i^T\Sigma^{-1}x_i}\right) - \sum_{i=1}^{N} c_{i,y}\log\left(r_y^d\sqrt{|2\pi\Sigma|}\right)\right)$$

נגזור לפי r_y ונשווה לאפס:

$$\frac{\partial E(l(S,\theta))}{\partial r_y} = \frac{\partial}{\partial r_y}\left(-\sum_{i=1}^{N}\frac{c_{i,y}}{2r_y^2}\cdot x_i^T\Sigma^{-1}x_i\right) - \sum_{i=1}^{N}\frac{c_{i,y}\, d\sqrt{|2\pi\Sigma|}\, r_y^{d-1}}{\sqrt{|2\pi\Sigma|}\, r_y^d} = 0$$

$$\Rightarrow -\sum_{i=1}^{N} \frac{-\cancel{4} r_y c_{iy}}{\cancel{4} r_y^{\cancel{4}\,3}} \cdot x_i^T \Sigma^{-1} x_i = \frac{r_y^{d-1}}{r_y^{d}} \sum_{i=1}^{N} d\, c_{iy}$$

$$\Rightarrow \frac{1}{r_y^3} \sum_{i=1}^{N} c_{iy} x_i^T \Sigma^{-1} x_i = \frac{d}{r_y} \sum_{i=1}^{N} c_{iy}$$

$$\Rightarrow \boxed{r_y^2 = \frac{\sum_{i=1}^{N} c_{iy} x_i^T \Sigma^{-1} x_i}{d \sum_{i=1}^{N} c_{iy}}}$$

Scanned by CamScanner