Երևանի Պետական Համալսարան

Մաթեմատիկայի և Մեխանիկայի ֆակուլտետ

Տվյալների գիտության և վիճակագրության բաժին

# Միջինացում-Մաքսիմալացում Ալգորիթմ

Հիմնական գաղափարներ, մաթեմատիկական հիմունքներ

Ուսանողներ: Լևոն Պապիկյան , Լուիզա Մուրադյան, Ռոբերտ Զաքոյան

1. Ներածություն ..... 3
2. Միջինացում-Մաքսիմալացում ալգորիթմի ընդհանուր դեպքում.....3
3. Գաուսյան խառնուրդ մոդել և Միջինացում-Մաքսիմալացում …….5
4. K-Means.....10
5. Գաուսյան խառնուրդ մոդելի կիրառումը MNIST տվյալների վրա……12

## Ներածություն

Վիճակագրական գնահատումը ծառայում է որպես տարբեր ոլորտներում տվյալների վրա հիմնված որոշումների կայացման գործիք՝ սկսած գիտական հետազոտություններից մինչև բիզնեսի վերլուծություն: Այնուամենայնիվ, բացակայող կամ թաքնված փոփոխականների առկայությունը զգալի դժվարություններ է ստեղծում մոդելի պարամետրերի ճշգրիտ գնահատման և տվյալների մեկնաբանման համար: Ավանդական գնահատման մեթոդները հաճախ չեն կարողանում արդյունավետ կերպով լուծել նման բարդությունները: Նման դեպքերում Միջինացում-Մաքսիմալացում (Expectation-Maximization, EM) ալգորիթմը առաջարկում է վիճակագրական գնահատման այնպիսի մոտեցում, որը մասնակի դիտարկված տվյալների պայմաններում ցուցաբերում է լավ արդյունքներ:
EM ալգորիթմը, որն առաջին անգամ ներդրվել է Դեմփսթերի, Լայրդի և Ռուբինի կողմից 1977 թվականին, այդ ժամանակվանից դարձել է հզոր գործիք մեքենայական ուսուցման, վիճակագրության և հաշվողական կենսաբանության ոլորտներում:
EM ալգորիթմը գործում է դիտարկվող տվյալների ճշմարտանմանության ֆունկցիան մաքսիմալացնելու սկզբունքով՝ կրկնվող կերպով գնահատելով թաքնված փոփոխականների և մոդելի պարամետրերի արժեքները: Այս կրկնվող գործընթացը ներառում է երկու հիմնական քայլ՝ Միջինացում (E) քայլ և Մաքսիմալացում (M): Այս քայլերի միջև հերթափոխելով՝ EM-ն աստիճանաբար ճշգրտում է իր գնահատումները մինչև ճշմարտանմանության ֆունկցիայի մաքսիմալացումը:
Թեև EM-ը կոնցեպտուալ առումով պարզ է, կան գործնական մարտահրավերներ՝ օրինակ՝ մեծ տվյալների ծավալները կամ չափողականությունը: Այս մարտահրավերների լուծումը հաճախ ներառում է այնպիսի մեթոդների կիրառումը, ինչպիսիք են մի քանի տարբեր սկզբնական արժեքներով վերագործարկումը կամ մեծ չափողականության դեպքում, դրա նվազեցումը:
Թեև EM-ն առաջարկում է բազմաթիվ առավելություններ բացակայող տվյալների մշակման և պարամետրերի գնահատման հարցում, այլընտրանքային մեթոդները, օրինակ՝ Մարկովյան շղթայի Մոնտե Կառլոյի (MCMC) տեխնիկան, առաջարկում են վիճակագրական գնահատման լրացուցիչ մոտեցումներ և կարող են գերազանցել EM-ին որոշակի սցենարներում:
Այս զեկույցում փորձել ենք մանրամասն ներկայացնել ալգորիթմը, մաթեմատիկական հիմունքները, ինչպես նաև ներառել ենք կիրառությունը R և Python ծրագրավորման լեզուներով:Աշխատանքում հիմնականում ներկայացված են ալգորիթմի տեսական գաղափարները:
Առաջին մասում բերվում է հիմնական գաղափարը, 2-րդ մասում՝ Գաուսյան խառնուրդ մոդելի օրինակ, որտեղ կիրառվում է Միջինացում-Մաքսիմալացում ալգորիթմը, 3-րդ մասում Գաուսյան խառնուրդ մոդելում կիրառված EM ալգորիթմի Մաքսիմալացում քայլից դուրս է բերվում K-Means ալգորիթմը,իսկ 4-րդ մասում բերվում է ալգորիթմի կիրառումը MNIST տվյալների վրա:

## Միջինացում-Մաքսիմալացում ալգորիթմը ընդհանուր դեպքում

Միջինացում-Մաքսիմալացում ալգորիթմը սովորաբար նկարագրվում է որպես երկու քայլ (E-քայլ և M-քայլ), բայց մենք կփոձենք այն բաժանել չորս քայլի. Դիցուք տրված է համատեղ բաշխում $p(X, Z \mid \theta)$, $X = (X_1, \ldots, X_N)$ պատահական նմուշի և $Z = (Z_1, \ldots, Z_M)$ անհայտ փոփոխականների համար, որոնք ևս պատահական մեծություններ են: Նպատակն է գտնել $p(X|\theta)$ ճշմարտանմանության ֆունկցիայի մաքսիմումը:

$$\ln p(X|\,\theta) \;=\; \ln[\sum_{Z} p(X, Z|\,\theta\,)]$$

Նկատենք, որ նույնիսկ այն դեպքում երբ, համատեղ բաշխման ֆունկցիան էքսպոնենտական ընտանիքից է, գումարի պատճառով $\ln p(X|\theta)$-ն չի լինի էքսպոնենտական ընտանիքից, որը բավականին բարդացնում է լոգարիթմական-ճշմարտանմանության ֆունկցիաի ուղղակի մաքսիմալացումը:
Այդ իսկ պատճառով կդիտարկենք Միջինացում-Մաքսիմալացում ալգորիթմը:

1. ընտրենք նախնական արժեքներ $\theta^{\text{old}}$

2. Հաշվի առնելով դիտարկված տվյալները՝ $x$, և կիրառելով մեր նախնական արժեք՝ $\theta^{\text{old}}$-ը, դիտարկենք պայմանական բաշխումը՝ $p(Z|x,\theta^{\text{old}})$

3. Օգտագործելով 2-րդ քայլում հաշվարկված $p(Z|x,\theta^{\text{old}})$ պայմանական բաշխումը, ձևավորենք պայմանական մաթեմատիկական սպասումը, որը կոչվում է Q-ֆունկցիա.

$$Q(\theta|\theta^{\text{old}}) = E(\ln p(X,Z|\theta) \mid x, \theta^{\text{old}}) = \sum_Z p(Z|x,\theta^{\text{old}}) \ln p(X,Z|\theta)$$

   Գտնենք այն $\theta$-ն, որը առավելագույնի է հասցնում Q-ֆունկցիան, արդյունքը մեր նոր գնահատականն է $\theta^{\text{new}}$՝

$$\theta^{\text{new}} = \operatorname{argmax}_\theta Q(\theta|\theta^{old})$$

4. Ստուգենք, արդյոք ունենք զուգամիտություն ճշմարտանմանության ֆունկցիայի կամ պարամետրերի արժեքների համար: Եթե զուգամիտության պայմանները բավարարված չեն, ապա՝

$$\theta^{\text{old}} \leftarrow \theta^{\text{new}}$$

   և վերադառնանք քայլ 2:

Ալգորիթմը չի սահմանում դադարեցման պայմանները: Մեր կողմից տրված պայմանները պետք է կրկնվեն այնքան ժամանակ, մինչև դրանք կբավարարվեն:
Օրինակ՝ $|l(\theta^{old}) - l(\theta^{\text{new}})| < \epsilon$ կամ $|\theta^{old} - \theta^{\text{new}}| < \epsilon$, ֆիքսված $\epsilon > 0$:
Միջինացում-Մաքսիմալացում գնահատականը միայն երաշխավորված է, որ երբեք չի վատանա:
Սովորաբար, այն կգտնի $p(X|\theta)$ ճշմարտանմանության ֆունկցիայի մաքսիմումը, բայց եթե $p(X|\theta)$ ճշմարտանմանության ֆունկցիան ունի մի քանի լոկալ մաքսիմումներ, Միջինացում-Մաքսիմալացումը կարող է չգտնել գլոբալ մաքսիմում արժեքը: Գործնականում կարելի է ալգորիթմը սկսել մի քանի պատահական սկզբնական արժեքներով անհայտ պարամետրերի համար և ընտրել մեծագույն արժեք ունեցող ճշմարտանմանության ֆունկցիայով գնահատականները որպես վերջնական գուշակություն $\theta$-ի համար:
Միջինացում-Մաքսիմալացում ալգորիթմի ավանդական նկարագրությունը բաղկացած է ընդամենը երկու քայլից.

1. **E- քայլ** Հաշվի առնելով նախորդ կրկնությունից $\theta^{old}$ գնահատականը, հաշվարկենք պայմանական մաթեմատիկական սպասումը $Q(\theta|\theta^{old})$

2. **M-քայլ** $\theta^{new}$ հետևյալն է.

$$\theta^{new} = \operatorname{argmax}_\theta Q(\theta|\theta^{old})$$

## Գաուսյան խառնուրդ մոդել և Միջինացում-Մաքսիմալացում

Այս բաժնում մենք կքննարկենք պարզ խառնուրդ մոդել (mixture model)՝ խտության ֆունկցիայի գնահատականի համար, և կից Միջինացում-Մաքսիմալացում ալգորիթմը՝ ճշմարտանմանության ֆունկցիան մաքսիմիզացնելու համար:
Նախ եկեք գեներացնենք տվյալներ R լեզվով: Դիտարկենք 0.2 հավանականությամբ բինոմական բաշխված $C_1 \ldots\ldots\ldots C_{10000}$ պատահական մեծությունների հաջորդականություն:

$$C_1, \ldots\ldots\ldots C_{10000} \sim Bin(0.2)$$

Դիտարկենք նաև մեկ այլ պատահական մեծությունների հաջորդականություն՝ $\beta_1 \ldots\ldots\ldots \beta_{10000}$,

$$\beta_i = \begin{cases} \eta \sim N(-12, 3^2), & C_i = 1 \\ \zeta \sim N(2, 1), & C_i = 0 \end{cases}$$

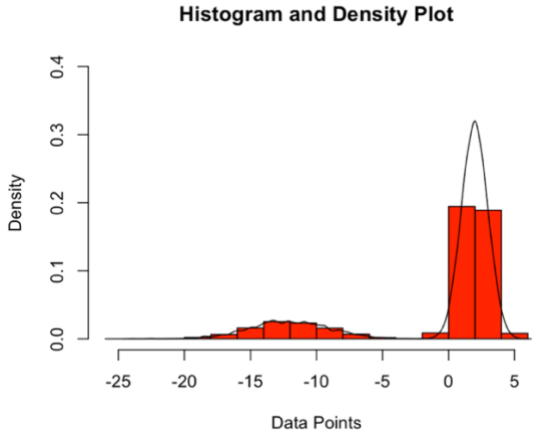
R լեզվով հաջորդաբար (համապատասխանաբար) գեներացնենք $C_i$ $\beta_i$ պատահական մեծությունների $c_1 \ldots\ldots\ldots c_{10000}$ , $y_1 \ldots\ldots\ldots y_{10000}$ թվային նմուշները:

```r
# Generate binomial samples
binomial_samples <- rbinom(10000, size = 1, prob = 0.2)


# Initialize an empty vector to store the generated values
datapoints <- numeric(length(binomial_samples))

# Generate values based on binomial samples
for (i in seq_along(binomial_samples)) {
  if (binomial_samples[i] == 1) {
    datapoints[i] <- rnorm(1, mean = -12, sd = 3)
  } else {
    datapoints[i] <- rnorm(1, mean = 2, sd = 1)
  }
}
```

Կազմելով նմանատիպ տվյալների խումբ, եկեք ենթադրենք, որ նորմալ բաշխումների միջինների ու վարիացիաների ինֆորմացիան կորցրել ենք:



Մենք ցանկանում ենք մոդելավորել մեզ տրված տվյալների խտությունը, և ակնհայտ երկմոդալության պատճառով Գաուսյան բաշխումը տեղին չի լինի: Գիտեք, որ հիմքում կան ընկած երկու առանձին ռեժիմներ, ուստի փոխարենը մենք կմոդելավորենք Y-ը որպես երկու նորմալ բաշխումների խառնուրդ.

$Y_1 \sim N(\mu_1, \sigma_1^2)$

$Y_2 \sim N(\mu_2,, \sigma_2^2)$

$Y = (1-\Delta)\cdot Y_1 + \Delta \cdot Y_2$

որտեղ $\Delta \in \{0,1\}$ $P(\Delta = 1) = p$ անհայտ հավանականությամբ: Y բաշխումից կամայական $Y_i$ պատահական մեծությանը կհամապատասխանի $\Delta_i$ պատահական մեծություն:
Հիմա ենթադրենք, որ մենք ցանկանում ենք առավելագույն հավանականությամբ այս մոդելը համապատասխանեցնել մեր տվյալներին: Պարամետրերն են՝

$$\theta = (p, \theta_1, \theta_2) = (p, \mu_1, \sigma^2{}_{1,}, \mu_2, \sigma^2{}_{2,})$$

N դեպքերի վրա հիմված լոգարիթմական-ճշմարտանմանության ֆունկցիան՝

$$L(\theta;\ X) = \sum_N log[(1-p)\varphi_{\theta_1}(y_i) + p\varphi_{\theta_2}(y_i)]$$

Որտեղ $\varphi_{\theta_1}(y)$ ,$\varphi_{\theta_2}(y)$ համապատասխանաբար $Y_1$, $Y_2$-ի բաշխման ֆունկցիաներն են: $L(\theta;\ X)$ -ի ուղղակի մաքսիմալացումը թվային առումով բավականին դժվար է, լոգարիթմի ներսում գումարի պատճառով : Սակայն, կա ավելի պարզ մոտեցում. Մենք հաշվի ենք առնում չդիտարկված, թաքնված (latent) փոփոխականները $\Delta_i$՝ դիտարկելով 0 կամ 1 արժեքները, քանի որ եթե $\Delta_i$ = 1, ապա $y_i$ -ն գալիս է 2-րդ մոդելից, հակառակ դեպքում այն գալիս է մոդել 1-ից: Ենթադրենք, որ մենք գիտեինք $\Delta_i$-ի արժեքները: Այդ դեպքում լոգարիթմական-ճշմարտանմանության ֆունկցիան կլինի հետևյալ տեսքի՝

$$\mathrm{L}(\theta;\ X, \Delta) = \sum_{k=1}^{N} l_0(\theta;\ X = y_k, \Delta_k) =$$

$$= \sum_{k=1}^{N}\left[(1-\Delta_k)\log\left(\phi_{\theta_1}(y_k)\right) + \Delta_k \log\left(\phi_{\theta_2}(y_k)\right)\right] + \sum_{k=1}^{N}[(1-\Delta_k)\log(1-\mathrm{p}) + \Delta_k \log(\mathrm{p})]$$

Եկեք նշանակենք $l_0(\theta;\ X = y_k, \Delta_k) := [(1-\Delta_k)\log(1-\mathrm{p}) + \Delta_k \log(\mathrm{p})] + [(1-\Delta_k)\log\left(\phi_{\theta_1}(y_k)\right) + \Delta_k \log\left(\phi_{\theta_2}(y_k)\right)]$ հետագա քայլերի համար:
$\mu_1, \sigma^2{}_1$-ի մաքսիմում ճշմարտանմանության գնահատականները կլինեն նմուշի՝ $\Delta_i$ = 0 խմբի, միջինը և դիսպերսիան, և նմանապես $\mu_2, \sigma^2{}_2$-ի համար գնահատականները կլինեն նմուշի միջինը և դիսպերսիան $\Delta_i$ = 1-ով տվյալների խմբի համար: p-ի Գնահատականը կլինի՝

$$\frac{\partial\, \mathrm{L}(\theta;\ X, \Delta)}{\partial \mathrm{p}} = \frac{-\sum_{i=1}^{N}(1-\Delta_i)}{1-\mathrm{p}} + \frac{\sum_{i=1}^{N}\Delta_i}{\mathrm{p}} = 0$$

$$\frac{1}{(1-\mathrm{p})\mathrm{p}}\sum_{i=1}^{N}\left((\Delta_i - 1)\mathrm{p} + \Delta_\mathrm{i}(1-\mathrm{p})\right) = 0$$

$$\hat{\mathrm{p}} = \frac{\sum_{i=1}^{N}\Delta_i}{N}$$

Քանի որ $\Delta_i$-ի արժեքներն իրականում անհայտ են, մենք կշարունակենք այլ եղանակով՝ փոխարինելով յուրաքանչյուր $\Delta_i$-ին իր մաթեմատիկական սպասման արժեքով:

$$\gamma_i(\theta) = E(\Delta_i|\theta, X) = \sum_{k=0}^{1} kP(\Delta_i = \mathrm{k}|\theta, X)$$

$$= 0P(\Delta_i = 0|\theta, X) + 1P(\Delta_i = 1|\theta, X) =$$

$$= 1P(\Delta_i = 1|\theta, X) = \frac{P(\Delta_i = 1)\, P(\theta, X\,|\Delta_i = 1)}{P(\theta, X)} = \frac{\mathrm{p}\varphi_{\theta_2}(\mathrm{y}_i)}{(1-p)\varphi_{\theta_1}(\mathrm{y}_i) + \mathrm{p}\varphi_{\theta_2}(\mathrm{y}_i)}$$

$$\gamma_i(\theta) = \frac{\mathrm{p}\varphi_{\theta_2}(\mathrm{y}_i)}{(1-p)\varphi_{\theta_1}(\mathrm{y}_i) + \mathrm{p}\varphi_{\theta_2}(\mathrm{y}_i)}$$

$\gamma_i(\theta)$ – ն կոչվում է նաև 2-րդ մոդելի պատասխանատվությունը $\mathrm{y}_i$ – ի համար. Մենք օգտագործում ենք EM ալգորիթմի ընթացակարգը, Գաուսի խառնուրդների հատուկ դեպքի համար: Միջինացում (Expectation) քայլում մենք յուրաքանչյուր $\mathrm{y}_i$ -ի համար կատարում ենք soft assignment, դիտարկում ենք $\mathrm{y}_i$ -ի 1-ին ու 2-րդ մոդելներին պատկանելու հավանականությունները: Մաքսիմալացման քայլում այս $\gamma_i(\theta)$ օգտագործվում են $Q(\theta, \theta^{old})$ ֆունկցիայի պարամետրերի գնահատումները թարմացնելու համար: $\mu_1^2$,$\mu_2^2$ -ի համար սկզբնական արժեքներ կառուցելու լավ միջոց է պարզապես պատահականորեն ընտրել երկուսը $\mathrm{y}_i$ -ից: Ինչպես $\sigma_1^2$ , այնպես էլ $\sigma_2^2$ -ի գնահատականները կարող են հավասար լինել թվային նմուշի ընդհանուր շեղմանը: p գնահատումը կարող է սկսվել 0,5 արժեքից:
Հիմա կառուցենք $Q(\theta, \theta^{old})$ ֆունկցիան մաքսիմալացման քայլի համար:

$$Q(\theta, \theta^{old}) = E\big(\mathrm{L}(\theta;\ X\ ,\Delta\ )|\theta^{old}, X = y)\big)$$

Որտեղ $\mathrm{y} = (\mathrm{y}_1, \ldots\ldots, \mathrm{y}_\mathrm{n})$ $\quad \theta^{old} = (p^{old}, {\theta_1}^{old}, {\theta_2}^{old})$

$$\mathrm{E}\big(\mathrm{L}(\theta;\ X\ ,\Delta\ )|\theta^{old}, X = y)\big) = \sum_{k=1}^{N} E\big(l_0(\theta;\ \mathrm{y_k}, \Delta_k)|\ \theta^{old}, X = \mathrm{y_k})\big) =$$

$$= \sum_{k=1}^{N} \sum_{\Delta_k = 0,1} P\ (\Delta_k|\ \theta^{old}, X = \mathrm{y_k}) l_0(\theta;\ \mathrm{y_k}, \Delta_k)$$

$$= \sum_{k=1}^{N} \sum_{\Delta_{k\,=\,0,1}} P(\Delta_k|\theta^{old}, X = \mathrm{y_k})\Big[\Big((1 - \Delta_k)\log\Big(\phi_{\theta_1}(y_k)\Big) + \Delta_k \log\Big(\phi_{\theta_2}(y_k)\Big)\Big)$$
$$+ \big((1 - \Delta_k)\log(1 - \mathrm{p}) + \Delta_k \log(\mathrm{p})\ \big)\Big] =$$

$$\sum_{k=1}^{N} \big(1 - \gamma_k(\theta^{old})\big)\Big(\log\Big(\phi_{\theta_1}(y_k)\Big) + \log(1 - \mathrm{p})\Big) + \sum_{k=1}^{N} \gamma_k(\theta^{old})\Big(\log\Big(\phi_{\theta_2}(y_k)\Big) + \log(\mathrm{p})\ \Big)$$

Հիմա դիտարկենք $Q(\theta, \theta^{old})$ ֆունկցիայի ածանցյալը ըստ $p$-ի

$$\frac{\partial Q}{\partial \mathrm{p}} = \sum_{k=1}^{N} \left(1-\gamma_k(\theta^{old})\right)\frac{-1}{1-\mathrm{p}} + \sum_{k=1}^{N} \gamma_k(\theta^{old})\frac{1}{\mathrm{p}} = 0$$

$$\frac{-\mathrm{p}\sum_{k=1}^{N}\left(1-\gamma_k(\theta^{old})\right) + (1-\mathrm{p})\sum_{k=1}^{N}\gamma_k(\theta^{old})}{(1-\mathrm{p})\mathrm{p}} = 0$$

$$-\mathrm{p}\sum_{k=1}^{N} 1 + \mathrm{p}\sum_{k=1}^{N}\gamma_k(\theta^{old}) + \sum_{k=1}^{N}\gamma_k(\theta^{old}) - \mathrm{p}\sum_{k=1}^{N}\gamma_k(\theta^{old}) = 0$$

$$\mathrm{pN} = \sum_{k=1}^{N}\gamma_k(\theta^{old})$$

$$\mathrm{p} = \frac{\sum_{k=1}^{N}\gamma_k(\theta^{old})}{N}$$

$p$ -ի գնահատականը ստացվեց $\hat{\mathrm{p}} = \frac{\sum_{k=1}^{N}\gamma_k\left(\theta^{old}\right)}{N}$

Դիտարկենք $Q(\theta,\theta^{old})$ ֆունկցիայի ածանցյալը ըստ $\mu_1$ - ի

$$\frac{\partial Q}{\partial \mu_1} = \sum_{k=1}^{N}\left(1-\gamma_k(\theta^{old})\right)\left(\frac{-(\mathrm{y_k}-\mu_1)}{\sigma_1^2}\right) = 0$$

$$\sum_{k=1}^{N}\left(1-\gamma_k(\theta^{old})\right)(\mathrm{y_k}-\mu_1) = 0$$

$$\sum_{k=1}^{N}\left(1-\gamma_k(\theta^{old})\right)\mathrm{y_k} - \sum_{k=1}^{N}\left(1-\gamma_k(\theta^{old})\right)\mu_1 = 0$$

$$\sum_{k=1}^{N}\left(1-\gamma_k(\theta^{old})\right)\mathrm{y_k} = \sum_{k=1}^{N}\left(1-\gamma_k(\theta^{old})\right)\mu_1$$

$$\mu_1 = \frac{\sum_{k=1}^{N}\left(1-\gamma_k(\theta^{old})\right)\mathrm{y_k}}{\sum_{k=1}^{N}\left(1-\gamma_k(\theta^{old})\right)}$$

$\mu_1$ -ի գնահատականը ստացվեց $\widehat{\mu_1} = \frac{\sum_{k=1}^{N}\left(1-\gamma_k(\theta^{old})\right)\mathrm{y_k}}{\sum_{k=1}^{N}\left(1-\gamma_k(\theta^{old})\right)}$

Համապատասխանաբար մյուս պարամետրերի գնահատականները կստացվի՝

$$\frac{\partial Q}{\partial \mu_2} = \widehat{\mu_2} = \frac{\sum_i^N \hat{\gamma}_i\, y_i}{\sum_i^N \hat{\gamma}_i}$$

$$\frac{\partial Q}{\partial \sigma^2{}_1} = \widehat{\sigma^2}_1 = \frac{\sum_i^N (1-\hat{\gamma}_i)(y_i-\widehat{\mu_1})^2}{\sum_i^N (1-\hat{\gamma}_i)} \qquad \frac{\partial Q}{\partial \sigma^2{}_2} = \widehat{\sigma^2}_2 = \frac{\sum_i^N \hat{\gamma}_i (y_i-\widehat{\mu_2})^2}{\sum_i^N \hat{\gamma}_i}$$

EM ալգորիթմը 2-կոմպոնենտով Գաուսյան խառնուրդ-ի համար.

1. Դիտարկել սկզբնական արժեքները $\widehat{\mu_1}$ , $\widehat{\mu_2}$ , $\widehat{\sigma^2}_1$, $\widehat{\sigma^2}_2$, $\hat{p}$
2. Միջինացման քայլ. Հաշվել պատասխանատվությունները

$$\hat{\gamma}_i(\theta) = \frac{\hat{p}\varphi_{\theta_2}(y_i)}{(1-\hat{p})\varphi_{\theta_1}(y_i)+\hat{p}\varphi_{\theta_2}(y_i)} \quad i = 1,2,\ldots,N$$

3. Մաքսիմալացման քայլ. Հաշվել կշռավորված միջինները ու վարիացաները:

$$\frac{\partial Q}{\partial \mu_1} = \widehat{\mu_1} = \frac{\sum_i^N (1-\hat{\gamma}_i) y_i}{\sum_i^N (1-\hat{\gamma}_i)} \qquad \frac{\partial Q}{\partial \mu_2} = \widehat{\mu_2} = \frac{\sum_i^N \hat{\gamma}_i y_i}{\sum_i^N \hat{\gamma}_i}$$

$$\frac{\partial Q}{\partial \sigma^2{}_1} = \widehat{\sigma^2}_1 = \frac{\sum_i^N (1-\hat{\gamma}_i)(y_i-\widehat{\mu_1})^2}{\sum_i^N (1-\hat{\gamma}_i)} \qquad \frac{\partial Q}{\partial \sigma^2{}_2} = \widehat{\sigma^2}_2 = \frac{\sum_i^N \hat{\gamma}_i (y_i-\widehat{\mu_2})^2}{\sum_i^N \hat{\gamma}_i}$$

և խառնուրդային հավանականությունը $\frac{\partial Q}{\partial p} = \hat{p} = \sum_i^N \hat{\gamma}_i / N$

4. կրկնել քայլ 2 ու 3 -ը մինչև զուգամիտություն

Մեր ունեցած տվյալների վրա R լեզվով Գաուսյան խառնուրդ մոդելը աշխատեցնելուց, ստացվեց հետևյալ արդյունքը՝

```
fitting ...
  |=================================================================| 100%
Iteration 1 : Mixing Proportions = 0.2014176 0.7985824
Iteration 1 : Mean Estimations = -12.07024 1.988329
Iteration 1 : Variance Estimations = 9.177795 0.9984295
fitting ...
  |=================================================================| 100%
Iteration 2 : Mixing Proportions = 0.2014175 0.7985825
Iteration 2 : Mean Estimations = -12.07025 1.988329
Iteration 2 : Variance Estimations = 9.177716 0.9984315
fitting ...
  |=================================================================| 100%
Iteration 3 : Mixing Proportions = 0.2014177 0.7985823
Iteration 3 : Mean Estimations = -12.07024 1.98833
Iteration 3 : Variance Estimations = 9.177817 0.998429
fitting ...
  |=================================================================| 100%
Iteration 4 : Mixing Proportions = 0.2014175 0.7985825
Iteration 4 : Mean Estimations = -12.07025 1.988329
Iteration 4 : Variance Estimations = 9.177721 0.9984314
fitting ...
  |=================================================================| 100%
Iteration 5 : Mixing Proportions = 0.2014177 0.7985823
Iteration 5 : Mean Estimations = -12.07024 1.98833
Iteration 5 : Variance Estimations = 9.177828 0.9984287
```

Վերջնական ճշմարտանմանության մաքսիմում մեթոդի գնահատականներն են՝
$\widehat{\mu_1} = -12.07024$ $\widehat{\mu_2} = 1.98833$ $\widehat{\sigma^2}_1 = 9.177828$ $\widehat{\sigma^2}_2 = 0.9984287$ $\hat{p} = 0.2014176$

## K-Means

K-means ալգորիթմի ու Գաուսի խառնուրդների EM ալգորիթմի միջև կա մոտ նմանություն: Մինչ K-means-ը կատարում է տվյալների կետերի hard assigment կլաստերներին՝ կետին համապատասխանեցնելով մեկ կլաստեր, EM ալգորիթմը ապահովում է soft assignment՝ կետին տալով յուրաքանչյուր կլաստերին պատկանելու հավանականություն:
Մենք կարող ենք դուրս բերել K-means ալգորիթմը որպես EM-ի հատուկ դեպք Գաուսի խառնուրդների համար: Դիտարկենք $\phi_{\theta_1}(y_k)$ , $\phi_{\theta_2}(y_k)$ բաշխման ֆունկցիաները դիտարկելով $\sigma_1^2 = \sigma_2^2 = \sigma^2$, որտեղ $\sigma^2$-ին ինչ-որ ֆիքսված հաստատուն է՝

$$\phi_{\theta_1}(y_k) = \frac{1}{\sqrt{(2\pi\sigma^2)}}\exp(-\frac{1}{2\sigma^2}|y_k - \mu_1|^2)$$
$$\phi_{\theta_2}(y_k) = \frac{1}{\sqrt{(2\pi\sigma^2)}}\exp(-\frac{1}{2\sigma^2}|y_k - \mu_2|^2)$$

$\sigma^2$ -ուն որպես հաստատուն վերաբերվելով, կետին տրված պատասխանատվություն $\gamma_i(\theta)$ -երն ընդունում են հետևյալ տեսքը՝

$$\gamma_i(\theta) = \frac{p\frac{1}{\sqrt{(2\pi\sigma^2)}}\exp(-\frac{1}{2\sigma^2}|y_k - \mu_2|^2)}{(1-p)\frac{1}{\sqrt{(2\pi\sigma^2)}}\exp(-\frac{1}{2\sigma^2}|y_k - \mu_1|^2) + p\frac{1}{\sqrt{(2\pi\sigma^2)}}\exp(-\frac{1}{2\sigma^2}|y_k - \mu_2|^2)}$$
$$= \frac{p\exp(-\frac{1}{2\sigma^2}|y_k - \mu_2|^2)}{(1-p)\exp(-\frac{1}{2\sigma^2}|y_k - \mu_1|^2) + p\exp(-\frac{1}{2\sigma^2}|y_k - \mu_2|^2)}$$

Եկեք դիտարկենք հետևյալ $r_i$ ֆունկցիան՝

$$r_i = \begin{cases} 1, & |y_i - \mu_2|^2 < |y_i - \mu_1|^2 \\ 0, & |y_i - \mu_2|^2 > |y_i - \mu_1|^2 \end{cases}$$

Այսինքն երբ $r_i$ ֆունկցիան ընդունում է 1 արժեքը, նշանակում է որ $y_i$ կետը ավելի մոտ է $\mu_2$ միջինին:

Ցույց տանք որ $\gamma_i(\theta)$ սահմանը երբ $\sigma^2 \to 0$ կստացվի հենց $r_i$: Դիտարկենք 2 դեպք՝

1. $|y_i - \mu_2|^2 < |y_i - \mu_1|^2$ նշանակենք $|y_k - \mu_2|^2 - |y_k - \mu_1|^2 := -c_1 \quad c_1 > 0$

$$\gamma_i(\theta) = \frac{p\exp\left(-\frac{1}{2\sigma^2}|y_k-\mu_2|^2\right)}{p\exp\left(-\frac{1}{2\sigma^2}|y_k-\mu_2|^2\right)\left(\frac{(1-p)\exp\left(-\frac{1}{2\sigma^2}|y_k-\mu_1|^2\right)}{p\exp\left(-\frac{1}{2\sigma^2}|y_k-\mu_2|^2\right)}+1\right)} = \frac{1}{\frac{(1-p)}{p}\exp(-\frac{1}{2\sigma^2}|y_k-\mu_1|^2+\frac{1}{2\sigma^2}|y_k-\mu_2|^2)+1}$$
$$= \frac{1}{\frac{(1-p)}{p}\exp(\frac{|y_k-\mu_2|^2-|y_k-\mu_1|^2}{2\sigma^2})+1} = \frac{1}{\frac{(1-p)}{p}\exp(\frac{-c_1}{2\sigma^2})+1}$$

$$\lim_{\sigma^2 \to 0} \frac{1}{\frac{(1-p)}{p}\exp(\frac{-c_1}{2\sigma^2})+1} = 1$$

2. $|y_i - \mu_2|^2 > |y_i - \mu_1|^2$ դեպքում քայլերը կատարվում է նույն տրամաբանությամբ և ստացվում՝

$$\lim_{\sigma^2 \to 0} \gamma_i(\theta) = 0$$

Այսպիսով, այս սահմանում մենք ստանում ենք տվյալների hard assignment կլաստերներին, Յուրաքանչյուր կետ վերագրվում է ամենամոտ միջինին ունեցող կլաստերին:

Դիտարկենք $\lim_{\sigma \to 0} \sigma^2 E\big(\mathrm{L}(\theta;\ X\ ,\Delta)|\theta^{old}, X = y)\big)$ սահմանը:

$$\lim_{\sigma \to 0} \sigma^2 E\big(\mathrm{L}(\theta;\ X\ ,\Delta)|\theta^{old}, X = y)\big) = \lim_{\sigma \to 0} \sigma^2 \left(\sum_{k=1}^{N}\big(1-\gamma_k(\theta^{old})\big)\left(\log\left(\frac{1}{\sqrt{(2\pi\sigma^2)}}\exp\left(-\frac{1}{2\sigma^2}|y_k-\mu_1|^2\right)\right) + \log(1-\mathrm{p})\right) + \sum_{k=1}^{N}\gamma_k(\theta^{old})\left(\log\left(\frac{1}{\sqrt{(2\pi\sigma^2)}}\exp\left(-\frac{1}{2\sigma^2}|y_k-\mu_2|^2\right)\right) + \log(\mathrm{p})\right)\right) = \lim_{\sigma \to 0} \sigma^2 \left(\sum_{k=1}^{N}\big(1-\gamma_k(\theta^{old})\big)\left(\log\left(\frac{1}{\sqrt{(2\pi\sigma^2)}}\right) - \frac{1}{2\sigma^2}|y_k-\mu_1|^2 + \log(1-\mathrm{p})\right) + \sum_{k=1}^{N}\gamma_k(\theta^{old})\left(\log\left(\frac{1}{\sqrt{(2\pi\sigma^2)}}\right) - \frac{1}{2\sigma^2}|y_k-\mu_2|^2 + \log(\mathrm{p})\right)\right) = \lim_{\sigma \to 0} \sigma^2 \left(\sum_{k=1}^{N}\big(1-\gamma_k(\theta^{old})\big)\log\left(\frac{1}{\sqrt{(2\pi\sigma^2)}}\right) - \sum_{k=1}^{N}\big(1-\gamma_k(\theta^{old})\big)\frac{1}{2\sigma^2}|y_k-\mu_1|^2 + \sum_{k=1}^{N}\big(1-\gamma_k(\theta^{old})\big)\log(1-\mathrm{p}) + \sum_{k=1}^{N}\gamma_k(\theta^{old})\log\left(\frac{1}{\sqrt{(2\pi\sigma^2)}}\right) - \sum_{k=1}^{N}\gamma_k(\theta^{old})\frac{1}{2\sigma^2}|y_k-\mu_2|^2 + \sum_{k=1}^{N}\gamma_k(\theta^{old})\log(\mathrm{p})\right) = \lim_{\sigma \to 0} \sigma^2 \left(\mathrm{N}\log\left(\frac{1}{\sqrt{(2\pi\sigma^2)}}\right) - \sum_{k=1}^{N}\big(1-\gamma_k(\theta^{old})\big)\frac{1}{2\sigma^2}|y_k-\mu_1|^2 + \sum_{k=1}^{N}\big(1-\gamma_k(\theta^{old})\big)\log(1-\mathrm{p}) - \sum_{k=1}^{N}\gamma_k(\theta^{old})\frac{1}{2\sigma^2}|y_k-\mu_2|^2 + \sum_{k=1}^{N}\gamma_k(\theta^{old})\log(\mathrm{p})\right) = -\frac{1}{2}\sum_{\mathrm{k}=1}^{\mathrm{N}}(1-r_k)|y_k-\mu_1|^2 - \frac{1}{2}\sum_{\mathrm{k}=1}^{\mathrm{N}} r_k|y_k-\mu_2|^2 + const$$

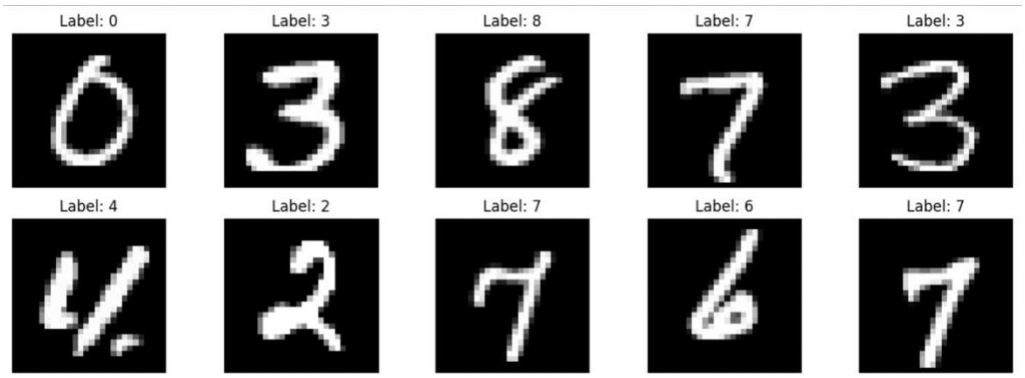
Այսինքն ստացվեց՝

$$\lim_{\sigma \to 0} \sigma^2 E\big(\mathrm{L}(\theta;\ X\ ,\Delta)|\theta^{old}, X = y)\big) = -\frac{1}{2}\sum_{\mathrm{k}=1}^{\mathrm{N}}(1-r_k)|x_k-\mu_2|^2 - \frac{1}{2}\sum_{\mathrm{k}=1}^{\mathrm{N}} r_k|x_k-\mu_2|^2 + const$$

Այսպիսով Գաուսյան խառնուրդ մոդելում EM ալգորիթմի Q ֆունկցիայի սահմանը՝ K-Means ալգորիթմում օգտագործվող, այն կարևոր ֆունկցիան է, որը չափում է կետերի և փնտրվող միջինների միջև հեռավորությունը ամեն իտերացիայում:

## Գաուսյան խառնուրդ մոդելի կիրառումը MNIST տվյալների վրա

Այս բաժնում կուսումնասիրենք MNIST տվյալների հավաքածուն և կփորձենք կիրառել Գաուսյան խառնուրդ մոդելը: MNIST տվյալների հավաքածուն բաղկացած է երկու մասից՝ ուսուցանող (train) հավաքածու, որը բաղկացած է 60,000 նկարներից և թեստային (test) հավաքածուն իր 10,000 նկարներով: Յուրաքանչյուր տող պարունակում է նկարի պիտակը (label) և pixel ինֆորմացիան՝ ընդհանուր 785 սյունով: Տվյալների հավաքածուն ներկայացնում է ձեռագիր թվերի պատկերների հավաքածու:




Սկզբում կդիտարկենք clustering-ի խնդիր: Այս գործընթացի ընթացքում մեր հիմական խնդիրներից մեկը տվյալների բարձր չափողականությունն է, ուստի կկատարենք չափողականության նվազեցում՝ օգտագործելով PCA: Քանի որ աշխատում ենք պատկերների հետ, տվյալների նորմավորումը (բաժանում 255-ի վրա) կարևոր քայլ է: Քանի որ ունենք 10 թվային դասեր, Գաուսյան խառնուրդ մոդելի կոմպոնենտների թիվը կսահմանենք 10: Նմանատիպ խմբավորման մոտեցումները կարող են օրինակ օգտակար լինել հասկանալու համար տվյալների մեջ եղած աղմուկը: 15 PCA կոմպոնենտով ստացված արդյունքում նկատում ենք որ, 7 և 9 թվերի զույգերը խմբաբորվել են 2-րդ և 5-րդ դասերում, ինչը նշանակում է, որ PCA կիրառելուց հետո դրանց մեջ կա բավականին նմանություն, իսկ մեր կառուցած GMM մոդելը նմանությունը չի կարողացել ճշտգրիտ տարբերակել: Մնացած թվերի դեպքում արդյունքը բավականին գոհացուցիչ է: Վերջնական ունեցանք 91.98% համընկնում:

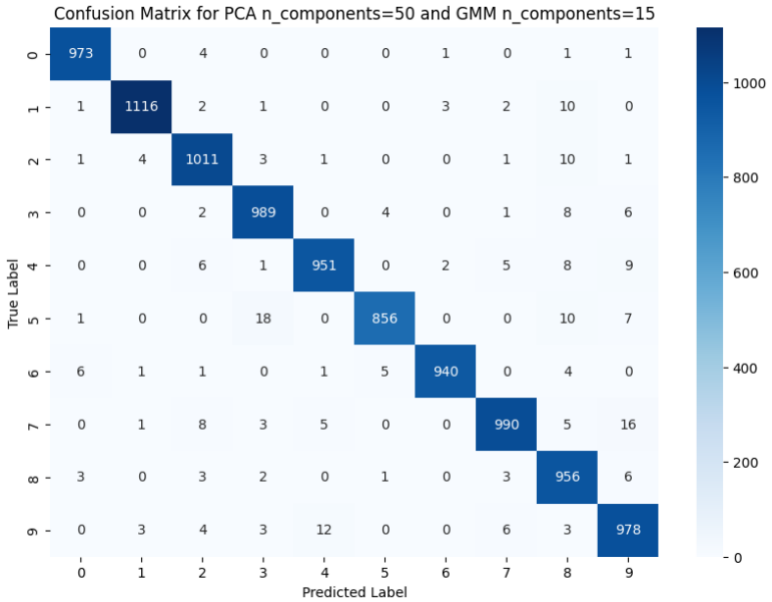
| cluster\label | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 6901 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 6982 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2194 | 9 | 3878 |
| 3 | 2 | 310 | 8 | 7141 | 5 | 2 | 0 | 0 | 0 | 0 |
| 4 | 0 | 7567 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 5099 | 6 | 3080 |
| 6 | 0 | 0 | 0 | 0 | 6818 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 6876 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 6311 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6810 | 0 |

Փորձենք նաև կառուցել մոդել, որը կկատարի classification տվյալների համար:
Կառուցենք 10 առանձին Գաուսյան խառնուրդ մոդելներ յուրաքանչյուր թվի համար: Նկատենք, որ այս մոտեցման ժամանակ label սյունը ուղիղ մասնակցություն չի ունենա ուսուցման ընթացքում: հետագայում անհայտ label -ով նկարը կպատկանի այն խմբին, որի մոդելը կտա ամենամեծ հավանականությունը: Նախքան ուսուցման գործընթացը, տվյալների մշակումը պարտադիր է: Այս մոտեցման շնորհիվ կարողացանք ստանալ ճշգրիտ կանխատեսումներ :

| PCA_n_components | GMM_n_components | Accuracy |
|---|---|---|
| 50 | 3 | 0.9698 |
| 50 | 5 | 0.9724 |
| 50 | 10 | 0.9771 |
| 50 | 15 | 0.9783 |
| 100 | 3 | 0.9583 |
| 100 | 5 | 0.9627 |
| 100 | 10 | 0.9648 |
| 100 | 15 | 0.9605 |
| 150 | 3 | 0.9408 |
| 150 | 5 | 0.9516 |
| 150 | 10 | 0.9513 |
| 150 | 15 | 0.9467 |

Այս աղյուսակը ներկայացնում է մեր ալգորիթմի տարբեր արդյունքները՝ PCA և GMM տարբեր բաղադրիչներով: Լավագույն արդյունքը ստացել ենք PCA-ի 50 և GMM-ի 15 բաղադրիչներով: Նկատում ենք, որ PCA-ի բաղադրիչների մեծացման հետ մոդելի կատարողականությունը վատանում է:

Լավագույն հիպերպարամետրերի համար գծագրել ենք confusion մատրիցը, և ընդհանուր արդյունքը բավականին գոհացուցիչ է:




Համեմատելու համար մեր մոդելի արդյունքները՝ փորձեցինք նաև ուսուցանել այլ մոդելներ, և ընդհանուր առմամբ՝ SVM և K-NN ալգորիթմները ունեն նույն ճշգրտությունը, ինչ մեր մոդելը:

Աղբյուրներ՝


• Pattern recognition and machine learning( M. bishop) - 423-448 էջեր

• The Elements of Statistical Learning:Data Mining,Inference and Prediction(Springer series in Statistics) -272-278 էջեր


• Հոդվածի բանաձևերի մանրամասն ստացումները,ինչպես նաև Գաուսյան խառնուրդների EM-ալգորիթմից K-means դուրս բերումը գրվել է՝ հիմվելով հավանականությունների տեսություն,մաթեմատիկական անալիզ դասընթացներից ունեցած գիտելիքների վրա:

Ինչպես նաև աշխատանքի կատարման ընթացքում հայտնաբերել ենք վրիպում Pattern recognition and machine learning( M. bishop) գրքում:

$$\lim_{\sigma\to 0} E\big(\mathrm{L}(\theta;\ X\ ,\Delta\ )|\theta^{old}, X = y)\big)\ = -\frac{1}{2}\sum_{\mathrm{k}=1}^{\mathrm{N}}(1-r_k)|x_k-\mu_2|^2 - \frac{1}{2}\sum_{\mathrm{k}=1}^{\mathrm{N}} r_k|x_k-\mu_2|^2\ + const$$

Փոխարենը պետք է լիներ՝

$$\lim_{\sigma\to 0}\sigma^2 E\big(\mathrm{L}(\theta;\ X\ ,\Delta\ )|\theta^{old}, X = y)\big)\ = -\frac{1}{2}\sum_{\mathrm{k}=1}^{\mathrm{N}}(1-r_k)|x_k-\mu_2|^2 - \frac{1}{2}\sum_{\mathrm{k}=1}^{\mathrm{N}} r_k|x_k-\mu_2|^2\ + const$$