EM - алгоритм

Видно? Слышно?

$\boxed{\text{EM - алгоритм.}}$

$\ell(y|\theta)$ трудно максимизируется.

например: наличие скрытых кластеров

модельная ситуация.

В.П. очень любит мёд

пчёлы $\to$ good (g) π_g

$\to$ bad (b) π_b

$\pi_g + \pi_b = 1$

тип пчёл: z_i

кол-во мёда: y_i

$$\boxed{(y_i | z_i = g) \sim N(\mu_g; 1)}$$

$$(y_i | z_i = b) \sim N(\mu_b; 1)$$

вер-сти $\forall$ (кроме одного) кластера.

$$\theta = (\underbrace{\mu_g, \mu_b}_{\text{пар-ры кластеров}}, \pi_g)$$

$\boxed{\text{Ситуация 1.}}$ В.П. умеет различ. z_i

день	мёд[кг] y_i	z_i (тип)
1	2	b
2	6	g
⋮	1	b
⋮	⋮	⋮
n	4	g

$y_i \in \mathbb{R} \Rightarrow$

$y_i \in \mathbb{R}^k$

[набл. незав]

i	y_i	z_i
1	2	b
2	4	g
3	1	b
4	6	g
5	2	b

$$\theta = (\mu_b, \mu_g, \pi_b)$$

$$\ln f(y,z|\theta) = \ln\Big[\pi_b \cdot f(y_1=2 | z_1=b, \theta) \cdot \pi_g \cdot f(y_2=4 | z_2=g, \theta) \cdot \ldots\ldots \Big]$$

$$\ln f(y,z|\theta) = \sum_{i=1}^{n} \ln\Big(I(z_i=b)\cdot \pi_b \cdot f(y_i | z_i=b, \theta) + I(z_i=g)\cdot \pi_g f(y_i | z_i=g, \theta)\Big)$$

иллюзия суммы из-за общего вида

good $\quad \pi_g$

bad $\quad \pi_b$

$$\pi_g + \pi_b = 1$$

опт. оч. легко !!

$$f(y_i | z_i=g, \theta) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{(y_i - \mu_g)^2}{2}\right)$$

в уме:

$$\hat{\mu}_g = \frac{\sum_{z_i=g} y_i}{\sum_{z_i=g} 1} = \bar{y}_{good}$$

$$\hat{\mu}_b = \frac{\sum_{z_i=b} y_i}{\sum_{z_i=b} 1}$$

$$\hat{\pi}_g = \frac{\sum_{z_i=g} 1}{n}$$

Ситуация 2

В.П. не раздирается в z_i

z_i - латентная пер-ая

i	y_i
1	•
2	⁞
⋮	⁞
n	⁞

z_i – не видим.

!!

$$\ell(y|\theta) = \sum_{i=1}^{n} \ln\left[\pi_б \cdot f(y_i|z_i=g,\theta) + \pi_в \cdot f(y_i|z_i=б,\theta)\right]$$

наст. сумма.

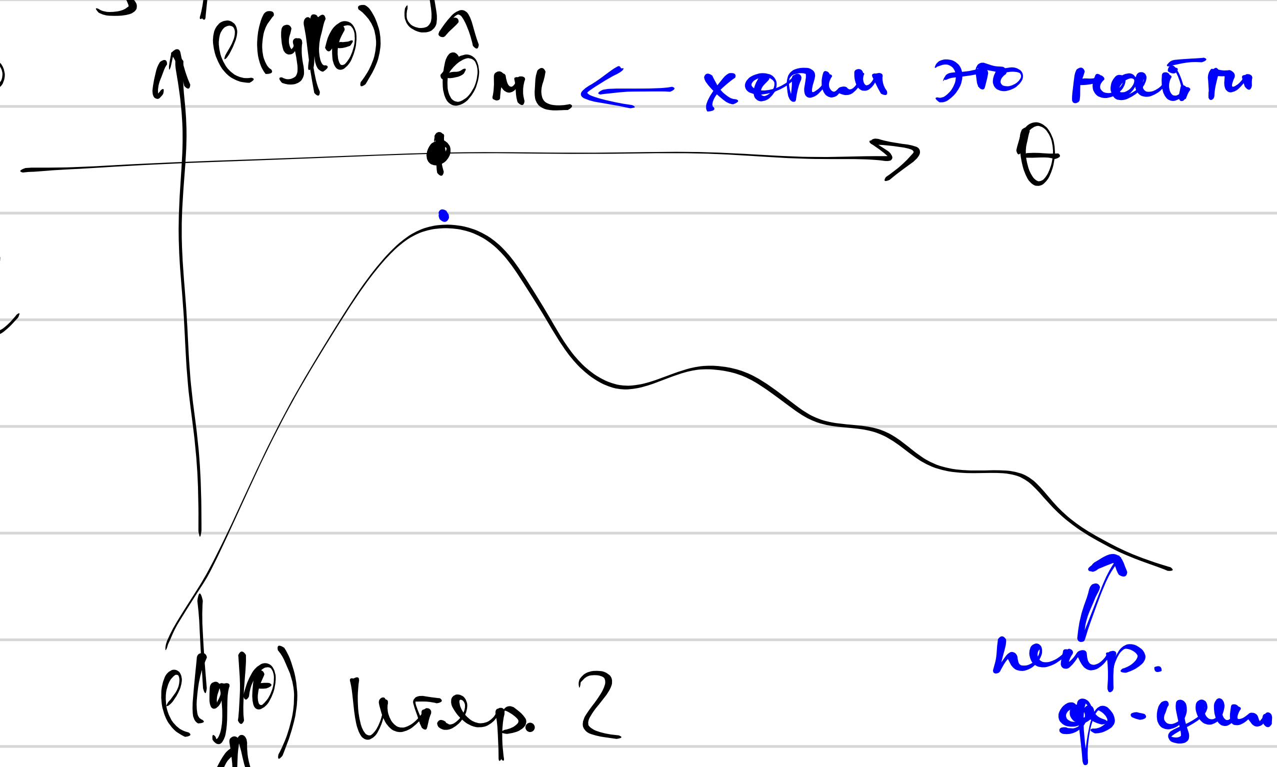
(не упр.-ся)

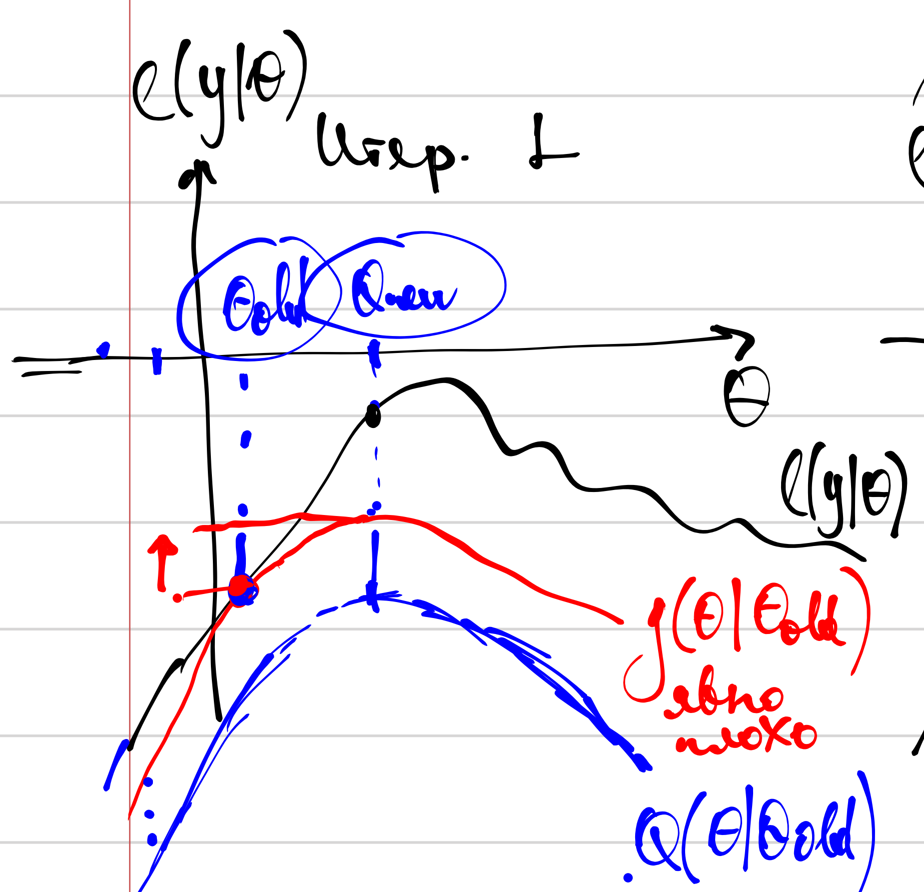
$$\boxed{\text{Цель: хотим } \max_{\theta} \ell(y|\theta)}$$

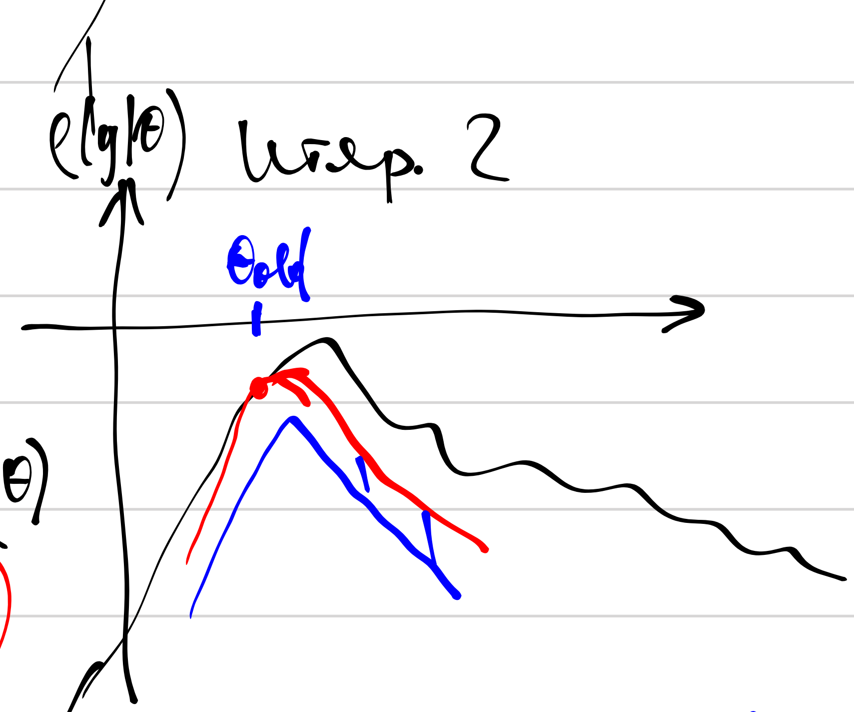
[к сожалению, станд. рассказ]

не теряя мотив-ию

$\boxed{\text{EM алгоритм в комиксах}}$!!







. . . . .

$$Q = g + const$$

$$\max_{\theta} Q(\theta|\theta_{old})$$

$$\ell(y|\theta) \geq g(\theta|\theta_{old})$$

$$\ell(y|\theta_{old}) = g(\theta_{old}|\theta_{old})$$

Иниц.

θ_{old} = [эвристика / от фонаря]

$$\boxed{\theta_{old} = (\mu_g = 5,\ \mu_b = ?,\ \pi_g = 0{,}1)}$$

E-шаг (expectation step)

E.1.

посчитаем: $p(z_i = clust \mid y_i, \theta_{old})$

i	y_i	z_i
1	2	b
2	4	g
3	1	b
4	6	g
5	2	b

↑ истина (не видим)

$$p(z_1 = g \mid y_1 = 2, \theta_{old}) = \boxed{\frac{P(A \cap B)}{P(B)}} = \frac{\pi_g \cdot f(y_1 = 2 \mid z_1 = g, \theta_{old})}{\sum_{clust} \pi_{clust} \, f(y_1 = 2 \mid z_1 = clust, \theta_{old})}$$

i	$p(z_i = g \mid y_i, \theta_{old})$	$p(z_i = b \mid y_i, \theta_{old})$
1	0,2	0,8
2	0,9	0.1
⋮		
5		

[что делать для непр. ой z_i? заморачиваться с нек-рой аппр-цией]

E.2. [здесь компьютер ничего не считает!] [это просто опр-ие ф-ции]

Init
E
M

$$Q(\theta \mid \theta_{old}) = E\left[\ell(y, z \mid \theta) \mid y, \theta_{old}\right]$$

E = expectation M = maximization

вар-ся
опр-ся
по θ_{old}

они взвешивают функции от θ.

Пример с ВП..

мы не считаем, а именно $\boxed{\text{определяем}}$ $Q(\theta|\theta_{old})$ как ф-цию от θ

$$Q(\theta|\theta_{old}) = E\left(l(y,z|\theta)\,\middle|\, y, \theta_{old}\right) =$$

$$= \sum_{i=1}^{n}\Big[p(z_i=g|y_i,\theta_{old})\cdot\ln\left(\pi_g\cdot f(y_i|z_i=g,\theta)\right) + p(z_i=b|y_i,\theta_{old})\cdot\ln\left(\pi_b\cdot f(y_i|z_i=b,\theta)\right)\Big]$$

предв. рассчит. конст на шаге E.1.

M-шаг.

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

Положим $\theta_{old} := \theta_{new}$ и пока не надоело переходим к E-шагу.

Init

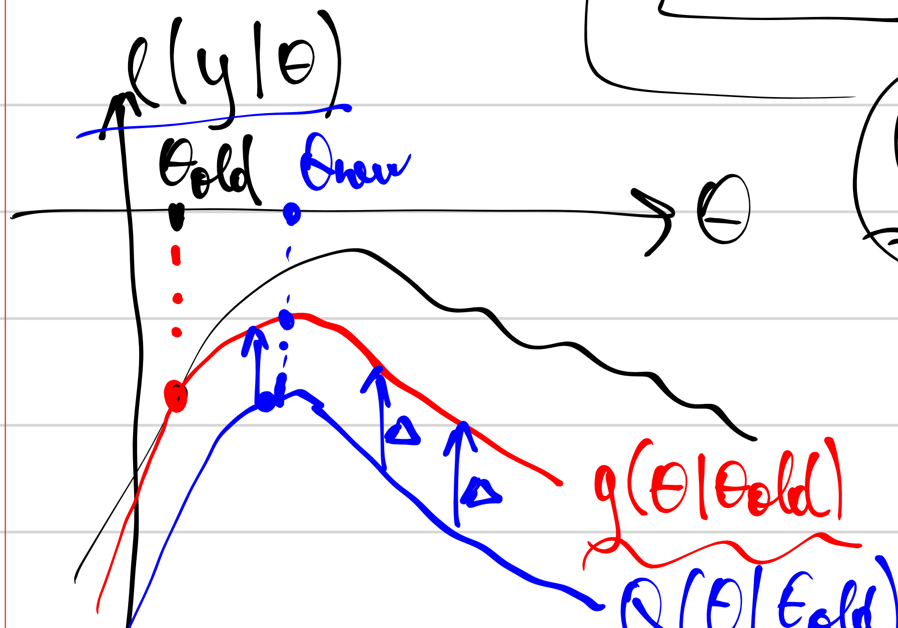
→ E-шаг E.1.
E.2

M-шаг

$\theta_{old} := \theta_{new}$

M!!

кас при $\theta=\theta_{old}$



$$l(y|\theta) = \ln\sum_z p(y,z|\theta) =$$

$$= \ln\sum_z \frac{p(y,z|\theta)}{p(z|y,\theta_{old})}\cdot p(z|y,\theta_{old}) =$$

$$= \ln E\left[\frac{p(y,z|\theta)}{p(z|y,\theta_{old})}\,\middle|\, y,\theta_{old}\right] \geq$$

$$\geq E\left[\ln p(y,z|\theta) - \ln p(z|y,\theta_{old})\,\middle|\, y,\theta_{old}\right]$$

$$= Q(\theta|\theta_{old}) - E\left[\ln p(z|y,\theta_{old})\,\middle|\, y,\theta_{old}\right]$$

→ $g(\theta|\theta_{old})$

$E(\ln\xi) \leq \ln E(\xi)$