AS in ML

Привет :)

Видно ли / слышно ли?

→ опишем EM на другом языке

→ Bootstrap: способ постр-ия довер. инт

EM алгоритм и KL-дивергенция.

задача: $\max\limits_{\theta} \ln p(x|\theta)$

заменяем на EM-алгоритм

Init. $\theta_{old} := \theta_{init}$

заменим на опт.

E-step. найдите $p(z|x,\theta)$ ← думать надо

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

M-step. $\max\limits_{\theta} Q(\theta, \theta_{old}) \to \theta_{new}$

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

Хотим поручить нахождение $p(z|x,\theta)$ компьютеру.

Сценарий

θ – неизв. параметр

x – наблюдения $\quad x = (x_1 \dots x_n)$

Скрытые z (латент) $\quad z = (z_1 \dots z_n)$

$$KL(q \| p_{z|x,\theta}) = CE(q \| p_{z|x,\theta}) - H(q)$$

распр-ие-кандидат

хотим найти

$$0 \le KL(a \| b) = \underbrace{CE(a \| b)}_{\text{среднее кол-во вопросов на разг-ие}} - \underbrace{H(a)}_{\text{среднее кол-во вопросов (в идеале)}}$$

лишние вопросы зад-ние разгад (в среднем)

a – ист-ое распр-ие загадывающего загадку.

b – распр-ие, кот-рое предп-т разгадывающий загадку.

!! Наивно: $\min\limits_{q} KL(q \| p_{z|x,\theta})$ $\quad q^* = p_{z|x,\theta}$

не сработает, т.к. чтобы вычислить KL надо уже знать $p_{z|x,\theta}$.

Лайфхак:

$$\underbrace{\ln p(x|\theta)}_{\text{не зав от } q} = \underbrace{KL(q \| p_{z|x,\theta})}_{\ge 0} + LB(q,\theta)$$

E-шаг: $LB(q,\theta) \to \max\limits_{q}$ $\quad \} \; q^* = p_{z|x,\theta}$

Чудо: для записи LB не нужно заранее знать $p_{z|x,\theta}$

[нат]

$$LB(q,\theta) = \ln p(x|\theta) - KL\left(q \,\|\, p_{z|x,\theta}\right) =$$

$$= \ln p(x|\theta) - \left[CE\left(q \,\|\, p_{z|x,\theta}\right) - H(q)\right] =$$

↑ распр-ие задающ. предп-ие отвод.
↑ ист-ное распр-ие данных

$$= \ln p(x|\theta) - \left[(-)\int q(z)\cdot \ln p(z|x,\theta)\,dz + \int q(z) \ln q(z)\, dz\right]$$

[бит] $CE(q\|p) = \int q(z) \log_{\frac{1}{2}} p(z)\, dz = -\int q(z) \log_2 p(z)\, dz$

[нат] $= -\int q(z) \ln p(z)\, dz \cdot \left[\frac{\text{бит}}{\text{нат}}\right]$

$$= \int q(z) \ln p(x|\theta)\, dz + \int q(z) \ln p(z|x,\theta)\, dz - \int q(z) \ln q\, dz$$

$$LB(q,\theta) = \int q(z) \cdot \ln \frac{p(x|\theta)\cdot p(z|x,\theta)}{q(z)}\, dz =$$

$$LB(q,\theta) = \int q(z) \cdot \ln \frac{p(x,z|\theta)}{q(z)}\, dz$$

$||$ ← [пропало $p(z|x,\theta)$]

EM-алгоритм

Init: $\theta_{old} := \theta_{init}$

E-шаг: $\max\limits_{q} LB(q, \theta_{old}) \longrightarrow q^{new}$ $\left[q^{new} = p(z|x,\theta)\right]$

M-шаг: $\max\limits_{\theta} LB(q^{new}, \theta) \longrightarrow \theta^{new}$

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

$$\ell(x|\theta_{old}) \overset{\text{E-шаг}}{=} LB(q^{new}, \theta_{old}) \overset{\text{M-шаг}}{\le} \underbrace{LB(q^{new}, \theta_{new})} \le$$

$$q^{new} = p(z|x, \theta_{old}) \qquad \le \ell(x|\theta_{new})$$

$$KL = 0 \qquad\qquad KL \ge 0$$

$$\ell(x|\theta_{new}) \ge \ell(x|\theta_{old}) \quad !!$$

E поиск $p(z|x,\theta)$ $\rightarrow$ E макс по q

M макс θ $\qquad$ M макс по θ

$$\max_\theta Q(\theta, \theta_{old}) \rightsquigarrow \max_\theta LB(q^{new}, \theta)$$

$\boxed{\text{Упр}}$ докажите эквив-сть.

Бутстрэп (Bootstrap)

— метод построения дов-х интервалов.

θ — неизв. пар-р
x — выборка.
$\hat\theta$ — оценка неизв-го θ
$\hat\theta = h(y)$

+ алгоритм $\hat\theta$ не важен
− алгоритм асимпт.

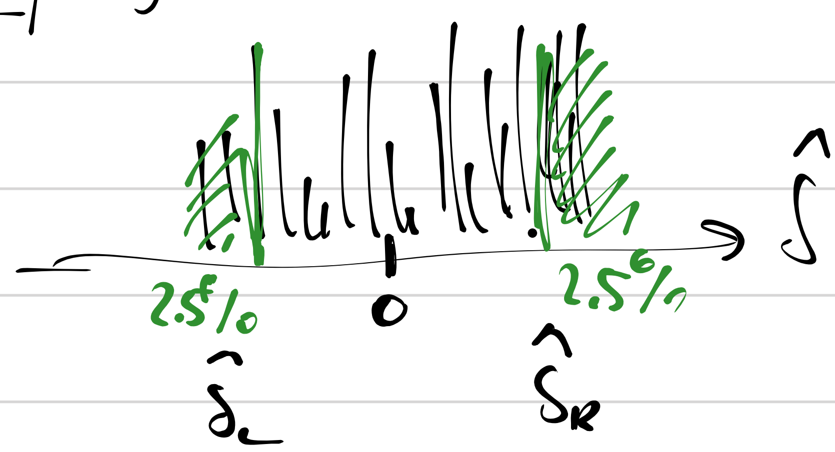
Эмпир-ий бутстрэп
$x \sim F$ F неизв

Парамет-ий бутстрэп.
$x \sim F_\theta$ F_θ известна с точностью до θ

→ reverse percentile bootstrap
→ t-bootstrap

$x_1 \ldots x_n \Rightarrow \hat\theta = h(y)$ $\hat\delta = \hat\theta - \theta$
не знаем!

$\boxed{\text{Если бы}}$ у меня была выборка $\hat\delta$



95%

$$\hat\delta_L \le \hat\theta - \theta \le \hat\delta_R \quad 95\%$$

$$-\hat\delta_R \le \theta - \hat\theta \le -\hat\delta_L$$

$$\theta \in [\hat\theta - \hat\delta_R ; \hat\theta - \hat\delta_L]$$

мечты!

Реально:

$$x_1, \dots x_n \to \hat{\theta} = h(x_1, \dots x_n)$$

$$\hat{\delta} = \hat{\theta} - \theta$$

одна штука — вообще не знаем

$$\boxed{\begin{gathered} \mu = E(x_i) \\ x_i \sim \text{iid с } E(x_i) < \infty \end{gathered}}$$

$\cancel{x_i \sim N(\mu; 1)}$

Бутстрэп-выборка

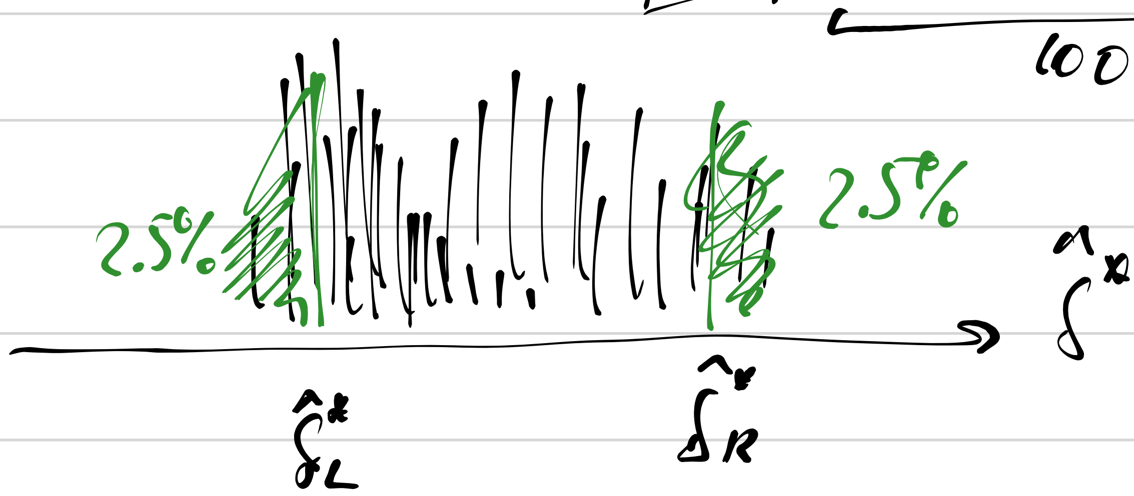
$$x_1^*, \dots x_n^* \to \hat{\theta}^* = h(x_1^* \dots x_n^*)$$

$$\hat{\delta}^* = \hat{\theta}^* - \hat{\theta} \leftarrow 10000 \text{ штук}$$

× 10000 раз

$$\begin{array}{ccccc} x & x^* & x^* & & x^* \\ 5 & 6 & 5 & & 6 \\ 6 & 6 & 5 & \dots & 2 \\ 2 & 2 & 5 & & 2 \end{array}$$

$\underbrace{\qquad\qquad\qquad}_{10000 \text{ раз}}$



$$CI_{Boot} = [\hat{\theta} - \hat{\delta}_R^* ; \hat{\theta} - \hat{\delta}_L^*]$$

Почему работает:

$x_i \sim F \leftarrow$ ист. ф. распр

$F^* \leftarrow$ выб. ф. распр-ия

n - велико $\Rightarrow \underline{F^* \approx F} \Rightarrow Law(\hat{\delta}^*) \approx Law(\hat{\delta})$

t-bootstrap.

есть теоремы, к-рые в идеальных условиях или при $n \gg 0$: $\frac{\hat\theta - \theta}{se(\hat\theta)} \to N(0;1)$

Если $x_1 \dots x_n \sim N(\mu;\sigma^2)$ независ

$\hat\mu = \bar y$ то $\frac{\bar y - \mu}{se(\bar y)} \sim t_{n-1}$

$se(\bar y) = \sqrt{\frac{\sum(y_i - \bar y)^2}{n-1}} = se$

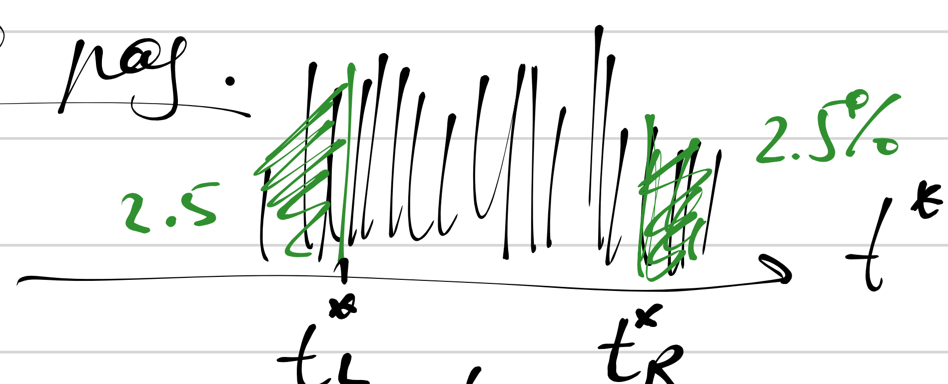
t-bootstrap:

$x_1 \dots x_n$

делаем бутстрэп выборку

$x_1^* \dots x_n^* \to \hat\theta^*$, $\to se^*$ $\qquad t^* = \frac{\hat\theta^* - \hat\theta}{se^*}$

× 10000 раз.



идеальный:

$CI = [\hat\theta - t_R \cdot se(\hat\theta); \hat\theta - t_L \cdot se(\hat\theta)]$

t-бутстрэп

$CI_{boot} = [\hat\theta - t_R^* \cdot se(\hat\theta); \hat\theta - t_L^* \cdot se(\hat\theta)]$

Парам. бутстрэп:

$x_1^* \dots x_n^* \sim$ генер случ $F_{\hat\theta}$