Bayesovské modelování

- využití: ekonometrie, medicína, sport, řízení, soudnictví, meteorologie...

$$P(H|D) = \frac{P(D|H)\cdot P(H)}{P(D)}$$

H – hypotéza (o parametru)
D – data

model (hustota, pravděpodobnostní funkce)

apriorní

marginální pravděpodobnost

$$= \int_H P(D|H)\cdot P(H)\, dH$$

$$\left(= \sum_H \sim \right)$$

$$\propto P(D|H)\cdot P(H)$$

až na normalizaci, aby $\int \cdot\, dH = 1$

aposteriorní pst

- vymyslet pár příkladů pro ilustraci

$H = \{$ pohlaví studenta, věk lektora, rychlost auta, hustota el. proudu, konverze ... $\}$

? $D|H$, D, ...

- regularizační role apriorna (extrémy – měření rovnováhy ve větru...)
- Připomenutí

$$P(A,B) = P(A|B)\cdot P(B)$$
$$= P(B|A)\cdot P(A)$$
$$= P(A)\cdot P(B) \quad \text{iff } A \perp\!\!\!\perp B$$

$$P(A) = \int_B P(A|B)\cdot P(B)\, dB$$
$$= \int_B P(A,B)\, dB$$

Formálně

x - data (vektor dat, ...) - např. měření spotřeby

θ - parametr (vektor) - např. střední hodnota spotřeby ...

α - hyperparametr

$\mathbb{X} = \{x_1, \dots, x_n\}$

$\tilde{x}$ - predikovaná hodnota

Model $\mathbb{X} \sim p(\mathbb{X}|\theta)$ - likelihood, věrohodnostní fce

Apriorní dist. $\theta \sim p(\theta|\alpha) = p(\theta)$ - α často nepíšeme a ani dále nebudeme (většinou)

Aposteriorní $p(\theta|\mathbb{X}) \propto p(\mathbb{X}|\theta)\, p(\theta)$

$$= \frac{p(X|\theta) \cdot p(\theta)}{\int p(X|\theta) \cdot p(\theta)\, d\theta} = \frac{p(X|\theta)\, p(\theta)}{p(X)}$$

↳ to znamená, že se změní α na α'

Predikce $p(\tilde{x}|\mathbb{X}) = \int p(\tilde{x}|\theta)\, p(\theta|\mathbb{X})\, d\theta$

↳ vyintegrujeme parametr

Jsou-li data nezávislá $p(x_1, \dots, x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$

↳ Bayes: $p(\theta|x_1, \dots, x_n) \propto p(\mathbb{X}|\theta)\, p(\theta)$

$$\propto p(\theta) \cdot \prod_{i=1}^{n} p(x_i|\theta)$$

[PF] **Hemofilie** – porucha srážlivosti krve
– vázaná na chromozom X
↪ ♂ XY – nemoc if X postižen
↪ ♀ XX – nemoc jen if oba X postiženy
⇒ přenašečka

Situace:

Žena ?

Její bratr – postižený (⇒ matka měla 1X postižený)

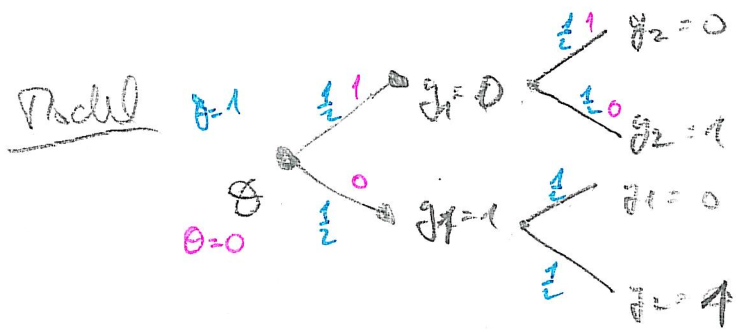
Její otec – nezasažený

$\Rightarrow \begin{cases} \theta = 1 & \text{(je nosič)} \\ \theta = 0 & \text{(není)} \end{cases}$

Apriorně $P(\theta=1) = P(\theta=0) = \frac{1}{2}$

Data: 2 synové $\begin{cases} y_1 \text{ – nezasažený} \\ y_2 \text{ – nezasažený} \end{cases}$

Model



$P(y_1=0, y_2=0 \mid \theta=1) = 0.5 \cdot 0.5 = 0{,}25$

$P(y_1=0, y_2=0 \mid \theta=0) = 1 \cdot 1 = 1$

Aposteriorně $P(\theta=1 \mid y)$?

$$P(\theta=1 \mid y) = \frac{p(y \mid \theta=1)\cdot p(\theta=1)}{p(y \mid \theta=1)\cdot p(\theta=1) + p(y \mid \theta=0)\cdot p(\theta=0)}$$

$$= \frac{\left(\frac{1}{2}\cdot\frac{1}{2}\right)\cdot\frac{1}{2}}{\left(\frac{1}{2}\cdot\frac{1}{2}\right)\cdot\frac{1}{2} + (1\cdot 1)\cdot\frac{1}{2}} = \frac{\frac{1}{8}}{\frac{5}{8}} = \underline{\underline{\frac{1}{5}}}$$

Sekvenční analýza – má 3. syna, nezasaženého. Uvažujeme předchozí aposteriorno jako apriorno → rekurzivní Bayes

$$p(\theta=1 \mid y_1, y_2, y_3) \propto p(y_3 \mid \theta)\cdot p(\theta \mid y_1, y_2)$$

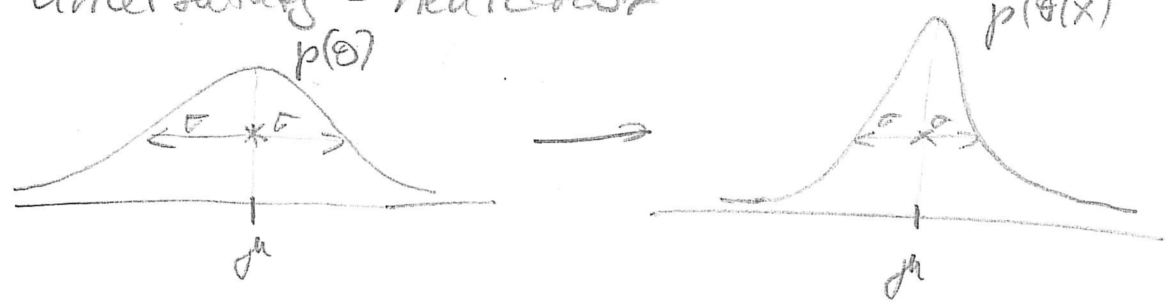
$$= \frac{\frac{1}{2}\cdot\frac{1}{5}}{\frac{1}{2}\cdot\frac{1}{5} + 1\cdot\frac{4}{5}} = \frac{\frac{1}{10}}{\frac{1}{10}+\frac{8}{10}} = \frac{1}{9} \quad (!)$$

Q: Co nám tedy říkal apriorní a aposť. pst?

A: Jaká je pst wrt počátek / konec experimentu

Q: A jak to bude vypadat ve spojitém světě?

↳ uncertainty = neurčitost



Přecházíme od pstí k distribucím

- připomenutí distribuce: p_i - psti !! jevů z pravděp. prostoru Ω

  $\sum_i p_i = 1$, $p_i \in [0,1]$, $p(\Omega) = 1$, $p(\emptyset) = 0$

spojitě: hustoty $p(\omega)$, $\omega \in \Omega \subset \mathbb{R}^n$

$$\int p(\omega)\, d\omega = 1$$

Když umíme pst pomocí míry, pak je nám spojitost jedno ☺

Apriorní distribuce

→ $p(\theta|X) \propto p(X|\theta)\, \underline{p(\theta)}$

- Teoreticky máme na $p(\theta)$ málo požadavků
  ↳ např. $\theta^{true} \in \operatorname{supp} p(\theta)$ → jinak by nešel odhadnout

- Prakticky chceme
  - → co nejlépe reprezentovat znalost (neurčitost)
    - → špičaté apriorno vs. ploché! (až úplně)
    - → neinformativní apriorno vlastně není ani h.p.!
  - → aby se nám dobře počítalo
    - → nepotřebujeme přesně fce, ale statistiky!?
    - → momenty, medián, modus... → z fce vypočteme

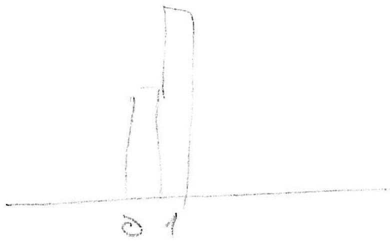
Př Ptáme se na férovost sítě.

Ping $\left\langle \begin{array}{l} 0 \text{ - paket se nevrátil} \\ 1 \text{ - vrátil} \end{array} \right\} x_i \in \{0,1\}$ → alternativní rozdělení (Bernoulliho)

? psť že se paket vrátil?

$p(x_i = 1) = \theta \in [0,1]$

$$f(X|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$$



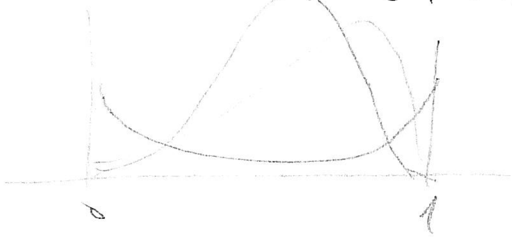
$$\boxed{\begin{aligned} \mathbb{E}[X] &= \theta \\ \operatorname{var} X &= \theta(1-\theta) \end{aligned}}$$

máme měření $X = \{x_1, x_2, \dots, x_n\} = \{0,1,1,0,1\}$

Apriorno

↳ beta

↳ $\theta \sim \beta(a,b)$



$$p(\theta) = \frac{1}{B(a,b)} \cdot \theta^{a-1}(1-\theta)^{b-1}$$

$$\mathbb{E}[\theta] = \frac{a}{a+b}$$

$$\operatorname{var} \theta = \frac{ab}{(a+b)^2(a+b+1)}$$

Bayes

$$\begin{aligned} p(\theta|X) &\propto p(X|\theta)\, p(\theta) \\ &\propto p(\theta) \cdot \prod_{i=1}^{n} p(x_i|\theta) \\ &\propto \frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1} \cdot \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1} \cdot \theta^{\Sigma x_i}(1-\theta)^{\Sigma(1-x_i)} \\ &= \frac{1}{B(a,b)} \theta^{(a+\Sigma x_i)-1}(1-\theta)^{(b+\Sigma(1-x_i))-1} \\ &\sim \beta\Big(a + \underbrace{\Sigma x_i}_{\#\text{úspěchů } x_i=1},\; b + \underbrace{\Sigma(1-x_i)}_{\#\text{neúspěchů } x_i=0}\Big) \end{aligned}$$

pokr.

↳ Čeho jsme dosáhli?

↳ apost. má stejný funkcionální tvar jako apriorní

Q: co když jí použijeme na nová x_n?

A: repete :)

Apriorní, které vedou na aposteriorní ze stejné třídy pstních rozdělení se nazývají **konjugovaná**

Tady bylo β = Alter × β

Norm. konstantu často nemusíme dopočítávat

Dokončení příkladu

↳ predikce

$$p(\tilde{x}=1|\mathbf{X}) = \int_0^1 p(\tilde{x}=1|\theta)\, p(\theta|\mathbf{x})\, d\theta$$

$$= \int_0^1 \theta\, p(\theta|\mathbf{x})\, d\theta$$

$$= \mathbb{E}[\theta|\mathbf{x}] = \frac{a'}{a'+b'}$$

Exponenciální třída, konjug. apriorna

$$ET: \quad p(x|\theta) = f(x) \cdot g(\theta) \cdot e^{\eta(\theta) \cdot T(x)}$$

fce

log-partition $e^{-\ln g(\theta) = A(\theta)}$

přir. par. (často $\eta(\theta) = \theta$)

suffic. stat. nez. na θ

$$\text{N.D.}: \prod p(x_i|\theta) = p(x_1, \dots, x_n|\theta) = f(x_1) \cdot \ldots \cdot f(x_n) \cdot (g(\theta))^n \cdot e^{\eta(\theta) \cdot \sum_{i=1}^{n} T(x_i)}$$

→ data jsou „obsažena" ve 2 statistikách ⇒

<u>více netřeba</u>. ξ, ν mají fixní rozměr!

$ET = \{\mathcal{N}, \chi^2_n, \text{Alter}, \text{Binomial}, \text{Poisson}, \text{Multi}, \dots\}$

Konjugovaná apriorna

$$p(\theta) = p(\theta|\nu, \xi) = h(\nu, \xi) \cdot [g(\theta)]^{\nu} \cdot e^{\eta(\theta)\,\xi}$$

⇒ <u>aposteriorna</u>:

$p(\theta|x) \propto p(x|\theta) \cdot p(\theta)$

$= \dots$

$\nu' = \nu + 1$

$\xi' = \xi + T(x)$

1 měření

⇒ <u>blokový (batch) update</u>:

$p(\theta|x_1, \dots, x_n) \propto p(\theta) \cdot \prod p(x_i|\theta)$

$= \dots$

$\nu' = \nu + n$

$\xi' = \xi + \sum T(x_i)$

<u>Př</u>: $\mathcal{N}(\mu, \cdot) \to \mu \sim \mathcal{N}(m, \tau^2)$

$\text{Alter}(\theta) \to \theta \sim \text{Beta}(a, b)$

$\text{Pois}(\lambda) \to \lambda \sim \text{Gamma}(a, b)$

$\text{Binom}() \to \beta(a, b)$

$\text{Multi}(p_1, \dots, p_k) \to \text{Dir}(\alpha_1, \dots, \alpha_k)$

PF | Poisson – Γ

Binom – β

$\mathcal{N}(\mu, \sigma^2)$ – $\mathcal{N}(\mu_0, \tau^2)$

PF | Lin. regrese

$$y_i = x_i \beta + e_i \qquad e_i \sim \mathcal{N}(0, \sigma^2)$$

známy

y_0, x_0 – pozorovatelné, $x_0 = [\cdot \; \cdot \; \cdot]$

$b = ?$ $\qquad \beta = [\;]$

$$y_i \sim \mathcal{N}(x_i \beta, \sigma^2)$$

$$p(y_0 | x_0, \beta, \sigma^2) \propto \exp\left\{ -\frac{1}{2\sigma^2} (y_0 - x_0 \beta)^2 \right\}$$

$$= \exp\left\{ -\frac{1}{2} \left[\sigma^{-2} \cancel{y_i^2} - 2 \cdot y_i \cdot \underbrace{x_i \beta}_{\text{skalár}} + (x_i \beta)(x_i \beta) \right] \right\}$$

$$\propto \exp\left\{ \mathrm{Tr} \left(\begin{bmatrix} -2\beta^T \\ \beta\beta^T \end{bmatrix} \right. \right.$$