Wyklad8

16 listopada 2012
00:23

Testowanie hipotez, Chernov, Sanov etc.

4. Testowanie Hipotez

4.1 Optymalne Bayesowskie testowanie hipotez

Obserwujemy realizacje zmiennej losowej X.
Zastanawiamy się czy pochodzi z rozkładu
$p(x|H_i)$ (hipoteza H_i) $i=1,\dots,M$
– tzw. dyskretna estymacja.

Zbiór wartości zmiennej losowej X : $\mathcal{X}$
dzielimy na $\mathcal{X}_i$ wprowadzamy
„funkcja decyzji"

$$h(x) = i \quad \text{dla } x \in \mathcal{X}_i$$

Jaka jest optymalna $h(x)$?

Przyjmijmy prawdopodobieństwa a priori:
$p(H_i)$, że wybrana została hipoteza H_i

Niech C_{ij} koszt wybrania hipotezy H_i jeśli prawdziwa była H_j. Np: $C_{ij} = 1 - \delta_{ij}$

Średni koszt dla danej funkcji decyzji:

$$C_h = \sum_j p(H_j) \int dx \, C_{h(x),j} \, p(x|H_j) =$$

$$= \int dx \sum_j C_{h(x),j} \, p(H_j|x) \, p(x)$$

Czyli powinniśmy wybrać:

$$\underbrace{h_B(x) = \arg\min_i \sum_j C_{ij}\, p(H_j|x)}$$

optymalna Bayesowska funkcja decyzji:

$$p(H_j|x) = \frac{p(x|H_j)\,p(H_j)}{p(x)} = \frac{p(x|H_j)\,p(H_j)}{\sum_j p(x|H_j)\,p(H_j)}$$

Dla $C_{ij} = 1 - \delta_{ij}$ (Minimum Error cost)

$$h_{ME}(x) = \arg\max_i p(H_i|x)$$

$$C_{ME} = 1 - \int dx \max_i [p(H_i|x)]\, p(x)$$

wybieramy i dla którego prawdopodobieństwo aposteriori największe. Równoważnie:

$$h_{ME}(x) = \arg\max_i p(x|H_i)\, p(H_i)$$

$$C_{ME} = 1 - \int dx \max_i (p(x|H_i)\, p(H_i)) =$$

Jeśli dodatkowo $p(H_i) = \frac{1}{M}$:

$$\arg\max_i p(H_i|x) = \arg\max_i \frac{p(x|H_i)\frac{1}{M}}{p(x)} =$$

$$= \arg\max_i p(x|H_i)$$

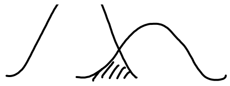
czyli wybieramy H_i dla której x najbardziej prawdopodobny: (Max-Likelihood).

$$h_{ML}(x) = \arg\max_i p(x|H_i)$$

$M = 2$

$p(H_0)\cdot p(x|H_0)$ $\quad$ $p(x|H_1)\cdot p(H_1)$

$C_{ij} = 1 - \delta_{ij}$



$$C_{ME} = 1 - \int dx \, \max\left(p(x|H_0)\, p(H_0),\ p(x|H_1)\, p(H_1) \right) =$$

$$= 1 - \frac{1}{2}\Big[\int dx \left| p(x|H_0)p(H_0) - p(x|H_1)p(H_1) \right|$$

$$+ \int dx\, p(x|H_0)p(H_0) + \int dx\, p(x|H_1)p(H_1) \Big) =$$

$$= \frac{1}{2}\Big[1 - \int dx \left| p(x|H_0)p(H_0) - p(x|H_1)p(H_1) \right| \Big]$$

$$= \frac{1}{2}\left[1 - \| p_1 - p_2 \| \right]$$

↑ total variation measure between prob. distributions

Przykład:

$X \overset{H_0}{\sim} N(\mu_0, \sigma_0^2)$ $p(H_0) = \pi_0$

$X \overset{H_1}{\sim} N(\mu_1, \sigma_1^2)$ $p(H_1) = \pi_1$ $C_{ij} = 1 - \delta_{ij}$

$$p(X|H_i) = \left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right)^N e^{-\frac{(X-\mu_i)^2}{2\sigma_i^2}}$$

$$p(x) = \sum_i p(x|H_i)\, p(H_i) = \sum_{i=0}^{1} \pi_i \left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right)^N e^{\frac{(x-\mu_i)^2}{2\sigma_i^2}}$$ – nie ważne

$$p(H_i|x) = \frac{\left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right)^N e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \cdot \pi_i}{p(x)}$$

Co jest większe $p(H_0|\bar{x})$ czy $p(H_1|\bar{x})$?

$$\frac{p(H_0|\vec{x})}{p(H_1|\vec{x})} = \frac{\pi_0}{\pi_1} \left(\frac{\sqrt{2\pi\sigma_1^2}}{\sqrt{2\pi\sigma_0^2}}\right) e^{-\frac{1}{2}\left(x^2\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) + (x)\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right) - \frac{N}{2}\left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_1^2}{\sigma_1^2}\right)\right)}$$

$$= \quad -\frac{1}{2}\left(x^2\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) - 2x\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right) + \frac{\mu_0^2}{\sigma_0^2} - \frac{\mu_1^2}{\sigma_1^2} \right)$$

$$= \frac{\pi_0}{\pi_1}\left(\frac{\sigma_1}{\sigma_0}\right) e$$

Jeśli ≥ 1 wybieramy H_0, jeśli < 1 wybieramy H_1

- Jeśli $\sigma_1 = \sigma_0$:

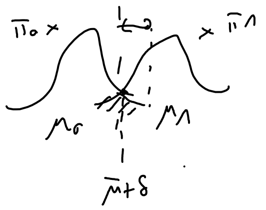
$$\frac{\pi_0}{\pi_1} e^{\frac{1}{\sigma^2}(\mu_0-\mu_1)x - \frac{(\mu_0^2-\mu_1^2)}{2\sigma^2}} \overset{H_0}{\gtrless} 1$$

$$\log\frac{\pi_0}{\pi_1} + \frac{1}{\sigma^2}(\mu_0-\mu_1)x + \frac{(\mu_0^2-\mu_1^2)}{2\sigma^2} \overset{H_0}{\gtrless} 0$$

$\{\mu_0 \leq \mu_1$

$$\langle x\rangle \overset{H_0}{\lesssim} \frac{\sigma^2\log\frac{\pi_0}{\pi_1} + \frac{\mu_0^2-\mu_1^2}{2}}{\mu_1-\mu_0} = \underbrace{\frac{\mu_0+\mu_1}{2}}_{\tilde{\mu}} + \underbrace{\frac{\sigma^2}{(\mu_1-\mu_0)}\log\frac{\pi_0}{\pi_1}}_{\delta}$$

Wtedy średni koszt błędu:



$$C = \pi_0 \int_{\tilde{\mu}+\delta}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu_0)^2}dx$$

$$+ \pi_1\int_{-\infty}^{\tilde{\mu}+\delta}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu_1)^2}dx$$

$$= \pi_0\frac{1}{\sqrt{\pi}}\int_{\frac{\mu_1-\mu_0}{2}+\delta}^{\infty} e^{-x^2}dx + \frac{\pi_1}{\sqrt{\pi}}\int_{-\infty}^{\frac{\mu_0-\mu_1}{2}+\delta} e^{-x^2}dx =$$

$$= \frac{\pi_0}{2}\left(1-\mathrm{Erf}\left(\delta+\frac{\mu_1-\mu_0}{2}\right)\right) + \frac{\pi_1}{2}\left(1-\mathrm{Erf}\left(\frac{\mu_1-\mu_0}{2}-\delta\right)\right)$$

4.2 Testowanie hipotez w granicy b. wielu kopii:

Przyjmijmy że chcemy rozróżnić pomiędzy dwoma hipotezami: H_0 i H_1

$$x \sim p(x|H_0) =: p_0(x), \quad \pi_0 = p(H_0)$$

$$x \sim p(x|H_1) =: p_1(x), \quad \pi_1 = p(H_1)$$

I jako koszt przyjmujemy $C_{ij} = 1 - \delta_{ij}$

Przyjmijmy że mamy N niezależnych realizacji x

czyli: $\vec{x} \sim \prod_i p_0(x_i)$ jeśli H_0

$\vec{x} \sim \prod_i p_1(x_i)$ jeśli H_1

Wiemy że powinniśmy wybrać H_0 jeśli

$$\frac{\pi_0 \prod_i p_0(x_i)}{\pi_1 \prod_i p_1(x_i)} \geq 1$$

I wtedy koszt będzie:

$$C_N = 1 - \int dx^{(N)} \max\left(\pi_0 \prod_i p_0(x_i), \pi_1 \prod_i p_1(x_i)\right)$$

W ogólności trudno obliczyć.

Chcemy znaleźć zachowanie asymptotyczne C_N

$$C_N \xrightarrow{N \to \infty} ?$$

Large deviation theory

$\cdot\ X \sim p(\cdot) \qquad x \in \mathcal{X} \qquad \mathcal{X} = \{a_1, a_2, \ldots a_{|\mathcal{X}|}\}$

N realizacji $x_1, \ldots, x_N =: \vec{x}$

Num $N(a|x^{(N)})$ - liczba obserwacji gdy $x_i = a$

Zdefiniujmy empiryczny rozkład prawdopodobieństwa

$$q_{\vec{x}}(a) = \frac{N(a|\vec{x})}{N}$$

↑ to samo w sobie jest zmienną losową.

Jakie jest prawdopodobieństwo, że zaobserwujemy dany empiryczny rozkład prawdopodobieństwa?

dany empiryczny rozkład prawdopodobieństwa ?

Jakie jest prawdop że zaobserwujemy konkretny ciąg $\vec{x}$:

$$p^{(N)}(\vec{x}) = \prod_i p(x_i) = \prod_a p(a)^{q_{\vec{x}}(a) \cdot N} =$$

$$2^{N \sum_a \log p(a) \cdot q_{\vec{x}}(a) \cdot + q_{\vec{x}}(a) \log q_{\vec{x}}(a) - q_{\vec{x}}(a) \log q_{\vec{x}}(a)}$$

$$= 2^{N \left(\sum_a q_{\vec{x}}(a) \log q_{\vec{x}}(a) - \sum_a q_{\vec{x}}(a) \log \frac{q_{\vec{x}}(a)}{p(a)} \right)}$$

$H(q) = -\sum_a q(a) \log q(a)$ — entropia Shanona

$D(q \| p) = \sum_a q(a) \log \frac{q(a)}{p(a)}$ — względna entropia

$$p^{(N)}(\vec{x}) = 2^{-N\left(H(q_{\vec{x}}) + D(q_{\vec{x}} \| p)\right)}$$

Ile jest ciągów N elementowych o tym samym rozkładzie empirycznym q

$$N_q^{(N)} = \frac{N!}{\prod_a (N q(a))!}$$

$$\log N_q^{(N)} = \log N! - \sum_a \log (N q(a))!$$

$$\log_2 N! = N \log_2 N - N \ln 2 + O(\ln N)$$

$$\frac{\log_a x}{\log_b x} = \log_b a$$

$$\ln x = \frac{\log_2 x}{\ln 2}$$

$$\frac{1}{N} \log N_q^{(N)} = \log N - \ln 2 \mp \sum_a (\log N + \log q(a)) q(a) + \ln 2\, q(a) + \frac{1}{N} O(\ln N)$$

$$= - \sum_a q(a) \log q(a) + \frac{1}{N} O(\log N)$$

$$N_q^{(N)} \doteq 2^{NH(q)}$$

Prawdopodobieństwo że zobaczymy jakiś ciąg o danym rozkł. empirycznym q

$$p^{(N)}(q) \doteq N_q^{(N)} \cdot 2^{-N(H(q) + D(q\|p))}$$

$$\boxed{p^{(N)}(q) \doteq 2^{-ND(q\|p)}}$$

interp. entr. względnej

Jeśli prawdziwy rozkład jest p jakie jest prawdop. że w N próbach uzyskamy empiryczny rozkład q

<u>Twierdzenie</u> Chernoff bound

$$C_N \doteq 2^{-N C(p_0, p_1)}$$

$$C(p_0,p_1) = -\log_2 \left[\min_{0 \le \lambda \le 1} \left(\int dx\, p_0(x)^{\lambda} p_1(x)^{1-\lambda} \right) \right] - \text{inf Chernoffa}$$

$\doteq$ równa się z dokładnością do pewnego nie-eksponencjalnego w N

Rozumiemy: $\lim_{N \to \infty} -\frac{1}{N} \log C_N = C(p_0, p_1)$

Dowód: (w wersji dyskretnej $\int dx \to \sum_{i=1}^{|X|}$)

Dowód: (w wersji dyskretnej $\int dx \to \sum_{i=1}^{|x|}$)

Wybieramy H_0 jeśli $\frac{\pi_0 \prod_i p_0(x_i)}{\pi_1 \prod_i p_1(x_i)} \geq 1$

równoważnie:

$$\log \frac{\pi_0}{\pi_1} + \sum_i \log \frac{p_0(x_i)}{p_1(x_i)} \geq 0$$

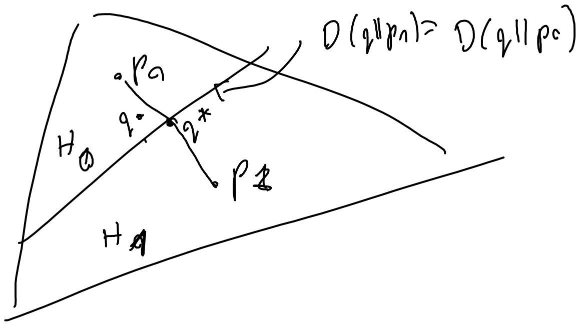
Niech $q(a) = \frac{N(a|\vec{x})}{N}$ – empiryczny rozkład $\vec{x}$

$$\log \frac{\pi_0}{\pi_1} + \sum_a N q(a) \log\left(\frac{p_0(a)}{p_1(a)} \cdot \frac{q(a)}{q(a)}\right) \geq 0$$

$$\log \frac{\pi_0}{\pi_1} + N\, D(q\|p_0) - N D(q\|p_1) \geq 0$$

dla dużych N:

$$D(q\|p_1) \leq D(q\|p_0)$$



Jeśli mieliśmy ciąg o empirycznym rozkładzie q w obszarze H_0 to popełnimy błąd z prawdop.

$$\doteq \pi_1 \cdot 2^{-N D(q\|p_1)}$$

Jeśli w obszarze H_1 to popełnimy błąd z prawd.

$$\doteq \pi_0\, 2^{-N(q\|p_0)}$$

Średni błąd:

Średni błąd:

$$C \doteq \pi_1 \sum_{q \in H_2} 2^{-ND(q|p_1)} + \pi_2 \sum_{q \in H_1} 2^{-ND(q|p_2)}$$

Rożnych typów jest nie więcej niż $(N+1)^{|X|}$ czyli dominujący wkład da czynnik o najmniejszym wykładniku. Dominujący wykładnik będzie równy:

$$D^* = \min_q D(q|p_1) \text{ z więzem}$$

$$D(q|p_1) = D(q|p_2)$$

i wtedy: $C \doteq 2^{-ND}$

Szukamy D^*: (mnożniki Lagrange'a)

$$l(q) = \sum_a q(a) \log \frac{q(a)}{p_1(a)} + \lambda \sum_a q(a) \log \frac{p_2(a)}{p_1(a)} + \nu \sum_a q(a)$$

$$\frac{dl}{dq(a)} = \log \frac{q(a)}{p_1(a)} + 1 + \lambda \log \frac{p_2(a)}{p_1(a)} + \nu = 0$$

$$2 \frac{q(a)}{p_1(a)} \cdot \left(\frac{p_2(a)}{p_1(a)}\right)^{\lambda} 2^{\nu} = 0$$

$$q(a) = \frac{1}{2} p_2(a)^{\lambda} p_1(a)^{1-\lambda} 2^{-\nu}$$

Czyli

$$q^*(a) = \frac{p_2(a)^{\lambda} p_1(a)^{1-\lambda}}{\underbrace{\sum_a p_2(a)^{\lambda} p_1(a)^{1-\lambda}}_{A_\lambda}}$$

t. że $D(q^*|p_1) = D(q^*|p_2)$

Czyli warunek na λ:

$$\sum_a \frac{p_2(a)^\lambda p_1(a)^{1-\lambda}}{A_\lambda} \log \frac{p_2(a)}{p_1(a)} = 0$$

$$\sum_a p_2(a)^\lambda p_1(a)^{1-\lambda} \log \frac{p_2(a)}{p_1(a)} = 0$$

$$\left\{ \begin{array}{l} \min\limits_\lambda \left(\sum_a p_2(a)^\lambda p_1(a)^{1-\lambda} \right) \\ \frac{d}{d\lambda} \left| \sum_a e^{\ln p_2(a) \cdot \lambda + \ln p_1(a)(1-\lambda)} \right| = \sum_a \ln \frac{p_2(a)}{p_1(a)} \, p_2(a)^{\lambda^*} p_1(a)^{1-\lambda^*} = 0 \end{array} \right.$$

$$D^* = \sum_a \frac{p_2(a)^{\lambda^*} p_1(a)^{1-\lambda^*}}{A_\lambda} \log \frac{p_2(a)^{\lambda^*} p_1(a)^{1-\lambda^*}}{A_\lambda \, p_1(a)} =$$

$$= -\sum_a \frac{p_2(a)^{\lambda^*} p_1(a)^{1-\lambda^*}}{A_\lambda} \log A_\lambda =$$

$$= -\log \sum_a p_2(a)^{\lambda^*} p_1(a)^{1-\lambda^*}$$

Czyli $\boxed{D^* = -\log \min\limits_\lambda p_2(a)^\lambda p_1(a)^{1-\lambda}}$