Cvičení z RPZ - Maximálně věrohodný odhad

Jan Šochman, Vojtěch Franc

23. března 2006

1 Úvod

Ve cvičení na Bayesovské rozhodování jsme využívali znalosti jak podmíněných pravděpodobností pozorování $p(x|y)$, tak apriorních pravděpodobností $P(y)$. V předchozím cvičení, jsme si ukázali minimaxní strategii pro případ, že neznáme apriorní pravděpodobnosti. Situace však může být ještě složitější a my nemusíme znát ani hustoty pravděpodobností $p(x|y)$. A právě jejich odhadem se budeme zabývat v tomto cvičení.

2 Maximálně věrohodný odhad

Nechť je hustota pravděpodobnosti $p(x|y, \theta)$ známa až na neznámou hodnotu parametru θ. V dalším výkladu se omezíme jen na jednu třídu (budeme hustoty pravděpodobností pro různé třídy odhadovat samostatně, předpokládáme tedy, že jsou nezávislé) a tak budeme používat jednodušší značení $p(x|\theta)$. Nechť je dále k dispozici trénovací množina

$$T = \{x_1, \ldots, x_n\}.$$

Prvky množiny T (měření, vzorky) byly získány náhodným výběrem z rozdělení $p(x|\theta)$. Důležitým předpokladem je, že měření musí být navzájem nezávislá. Naším cílem je na základě trénovací množiny T najít co nejvěrohodnější odhad parametru θ.

Správnost odhadu θ na základě trénovací množiny T budeme měřit pomocí podmíněné pravděpodobnosti $P(T|\theta)$, které budeme říkat *věrohodnost*. Díky nezávislosti prvků množiny můžeme snadno napsat pravděpodobnost trénovací množiny jako

$$P(T|\theta) = \prod_{i=1}^{n} p(x_i|\theta). \qquad (1)$$

Chceme najít takovou hodnotu parametru θ, která bude maximalizovat věrohodnost (odtud metoda maximální věrohodnosti). Hledáme tedy θ^*

$$\theta^* = \operatorname*{argmax}_\theta P(T|\theta) = \operatorname*{argmax}_\theta \prod_{i=1}^{n} p(x_i|\theta). \tag{2}$$

3 Nalezení maximálně věrohodného odhadu

K nalezení nejvěrohodnější hodnoty parametru θ využijeme nejprve skutečnosti, že pokud maximalizujeme nějakou funkci, její transformace pomocí rostoucí funkce zachová maximum. Jako transformační funkci si vybereme logaritmus (protože mění součiny na součty, což je pro počítání šikovné) a z rovnice (2) dostáváme

$$\theta^* = \operatorname*{argmax}_\theta \sum_{i=1}^{n} \log p(x_i|\theta) \tag{3}$$

Označme si maximalizovaný výraz

$$L(T, \theta) = \sum_{i=1}^{n} \log p(x_i|\theta). \tag{4}$$

Maximum L lze v některých případech (L musí být konvexní) získáme snadno zderivováním a položením derivace rovno nule

$$\frac{\partial L(T,\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \log p(x_i|\theta)}{\partial \theta} = 0. \tag{5}$$

4 Příklad – jednorozměrné normální rozdělení

Podívejme se blíže na příklad, kde hustota pravděpodobnosti má tvar normálního rozdělení, pro jednoduchost pouze jednorozměrného

$$p(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{6}$$

Předpokládejme, že známe směrodatnou odchylku σ, ale neznáme střední hodnotu μ. Máme pouze trénovací množinu $T = \{x_1, \ldots, x_n\}$. Hledáme tedy μ takové, že

$$\sum_{i=1}^{n} \frac{\partial \log p(x_i|\mu)}{\partial \mu} = 0. \tag{7}$$

Zjednodušením (logaritmus a exponenciála se vyruší) a zderivováním tohoto výrazu dostáváme

$$\sum_{i=1}^{n} \frac{(x_i - \mu^*)}{\sigma^2} = 0. \tag{8}$$

Tento výraz nabývá nulové hodnoty v případě, že

$$\sum_{i=1}^{n} (x_i - \mu^*) = 0. \tag{9}$$

Přeuspořádáním tohoto výrazu dostáváme výsledek

$$\mu^* = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{10}$$

Odvodili jsme maximálně věrohodný odhad střední hodnoty normálního rozdělení se známou směrodatnou odchylkou. Podobným způsobem si můžete vyzkoušet odvodit maximálně věrohodný odhad směrodatné odchylky, pokud známe střední hodnotu. Dojdete k výsledku

$$\sigma^* = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2} \tag{11}$$

5 Příklad – vícerozměrné normální rozdělení

Stejného postupu jako v jednorozměrném případě lze použít i pro odhad parametrů d-rozměrného normálního rozdělení

$$p(x|\mu, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det \Sigma}} e^{-\frac{1}{2}(x-\mu)^T \mathbf{\Sigma}^{-1} (x-\mu)} \tag{12}$$

Uvědomte si, že zde jsou měření x d-rozměrné vektory stejně tak jako střední hodnota μ. $\mathbf{\Sigma}$ je disperzní maticí rozdělení. Zde dojdeme k výsledkům

$$\mu^* = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{13}$$

$$\mathbf{\Sigma}^* = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T \tag{14}$$

6 Zadání úlohy

- Vygenerujte v Matlabu sekvenci vzorků pomocí funkce *gsamp* pro zvolené parametry normálního (Gaussovského) rozdělení, např. pro $\mu = 1$ a $\sigma^2 = 2$. Připravte sekvence o 5, 50, 500 vzorcích.

- Spočtěte maximálně věrohodný odhad střední hodnoty μ a směrodatné odchylky σ. Vykreslete si pro odhadnutou střední hodnotu věrohodnostní funkci jako funkci σ. Porovnejte přesnost odhadů a charakter průběhu věrohodnostní funkce v závislosti na počtu vzorků.

- Vygenerujte si 250 vzorků z dvourozměrného normálního rozdělení s parametry
$$\mu = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \mathbf{\Sigma} = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$
Zobrazte si vzorky (funkce *ppatterns*). Odhadněte střední hodnotu a disperzní matici. Do stejného grafu, ve kterém máte vzorky, zobrazte vrstevnice $p(\mathbf{x}|\mu, \mathbf{\Sigma})$ se skutečnými a odhadnutými parametry (použijte funkci *pgauss* nebo kombinaci *meshgrid, contour*).