

Zestaw 3 - Zadanie 7

Metody probabilistyczne w uczeniu maszynowym

Łukasz Trzos

Treść zadania

Dane wyjściowe y odpowiadające macierzy planowania X wygenerowane zostały z rozkładu $\mathcal{N}(X\theta, \sigma^2 \mathbf{I})$ dla pewnego $\sigma \neq 0$, gdzie θ ma rozkład a priori $\mathcal{N}(0, \tau^2 \mathbf{I})$ dla pewnego $\tau \neq 0$. Wyznacz wartość oczekiwaną rozkładu a posteriori $p(\theta|D)$. Znajdź związek między parametrem regularyzacji λ w problemie regresji grzbietowej oraz stałymi τ i σ .

Rozwiązanie

Wartość oczekiwana rozkładu a posteriori

Chcemy wyznaczyć $E[P(\theta|y)]$ Korzystamy z tw. Bayesa:

$$P(\theta|y) \propto P(y|X, \theta)P(\theta)$$

$P(y|X, \theta)$ ma rozkład normalny:

$$P(y|X, \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}(y-X\theta)^T(y-X\theta)}$$

$P(\theta)$ jest rozkładem a priori parametru θ :

$$P(\theta) = \frac{1}{(2\pi\tau^2)^{\frac{k}{2}}} e^{-\frac{1}{2\tau^2}\theta^T\theta}$$

Możemy więc pokazać, do czego proporcjonalny jest rozkład a posteriori względem θ :

$$\begin{aligned} P(\theta|y) &\propto \exp\left\{-\frac{1}{2\sigma^2}(y-X\theta)^T(y-X\theta) - \frac{1}{2\tau^2}\theta^T\theta\right\} = \exp\left\{-\frac{1}{2\sigma^2}y^Ty - \frac{1}{2\sigma^2}2y^TX\theta - \frac{1}{2\sigma^2}\theta^TX^TX\theta - \frac{1}{2\tau^2}\theta^T\theta\right\} \propto \\ &\propto \exp\left\{-\frac{1}{2}\theta^T\left(\frac{X^TX}{\sigma^2} + \frac{\mathbf{I}}{\tau^2}\right)\theta - \frac{1}{\sigma^2}y^TX\theta\right\} = \exp\left\{-\frac{1}{2}\theta^TA\theta - \frac{1}{\sigma^2}y^TX\theta\right\} \end{aligned}$$

dla $A = \frac{X^TX}{\sigma^2} + \frac{\mathbf{I}}{\tau^2}$. A jest macierzą dodatnie określoną, więc ma dodatni wyznacznik, a więc istnieje jej odwrotność.

Obliczmy jak wygląda rozkład prawdopodobieństwa dla wielowymiarowego rozkładu normalnego o wektorze wartości oczekiwanych $\mu = A^{-1}\frac{X^Ty}{\sigma^2}$ i macierzy kowariancji $\Sigma^{-1} = A$. Jest on proporcjonalny do:

$$\begin{aligned} &\propto \exp\left\{-\frac{1}{2}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)\right\} \propto \exp\left\{-\frac{1}{2}\theta^T\Sigma^{-1}\theta + \mu^T\Sigma^{-1}\theta\right\} = \exp\left\{-\frac{1}{2}\theta^TA\theta + \frac{1}{\sigma^2}(A^{-1}X^Ty)^T A\theta\right\} = \\ &= \exp\left\{-\frac{1}{2}\theta^TA\theta - \frac{1}{\sigma^2}y^TX\theta\right\} \end{aligned}$$

W pierwszej proporcji skorzystaliśmy z niezależności μ oraz Σ od parametru θ , co pozwala pominąć ostatni składnik, a w ostatniej równości wykorzystujemy fakt, że A jest macierzą symetryczną ze względu na swoją definicję, więc $(A^{-1})^T A = \mathbf{I}$.

Dochodzimy do wniosku, że rozkład prawdopodobieństwa $P(\theta|y)$ odpowiada rozkładowi normalnemu z odpowiednio dobranymi parametrami. Możemy więc obliczyć jego wartość oczekiwaną:

$$E[P(\theta|y)] = \mu = A^{-1}\frac{X^Ty}{\sigma^2} = \left(\frac{X^TX}{\sigma^2} + \frac{\mathbf{I}}{\tau^2}\right)^{-1}\frac{X^Ty}{\sigma^2} = (X^TX + \frac{\sigma^2}{\tau^2}\mathbf{I})^{-1}X^Ty$$

Związek z problemem regresji grzbietowej

Wiemy, że rozwiązanie problemu regresji grzbietowej z parametrem regularyzacji λ jest postaci:

$$\hat{\theta}^r = (X^T X + \lambda \mathbf{I})^{-1} X^T y$$

Przyjmując $\lambda = \frac{\sigma^2}{\tau^2}$, rozwiązania tych dwóch problemów mają taką samą postać.