

MAE5911/IME: Fundamentos de Estatística e Machine Learning. Prof.: Alexandre Galvão Patriota

Questão 01: Considere uma amostra aleatória $(Y_1, X_1), \dots, (Y_n, X_n)$ de (Y, X) tal que a distribuição condicional $Y | X = x \sim \mathcal{N}(\mu_\theta(x), \sigma_\theta^2(x))$, e suponha que a distribuição de X não contém informação sobre os parâmetros.

1.1 Apresente uma **rede neural** que modele o **quantil de ordem 75%** (terceiro quartil) da distribuição condicional $Y | X = x$. O código deve ser generalizável para qualquer quantil.

1.2 Mostre a aplicação do método nos seguintes dados simulados em R:

```
set.seed(32)
n = 1000
x = sort(runif(n, -4, 4))
y = 3/(3 + 2 * abs(x)^3) + exp(-x^2) + cos(x) * sin(x) + rnorm(n) * 0.3
```

Sugestão: reescreva a esperança e variância condicionais em termos do quantil e construa a função de verossimilhança apropriada.

1 Modelagem do quantil condicional via Normal

Considere o modelo condicional

$$Y | X = x \sim \mathcal{N}(\mu(x), \sigma^2(x)). \quad (1)$$

Para um nível de quantil $q \in (0, 1)$, definimos o quantil condicional $Q_q(x)$ de $Y | X = x$ como o valor t que satisfaz

$$\mathbb{P}(Y \leq t | X = x) = q. \quad (2)$$

No caso da Normal, o quantil condicional pode ser escrito em termos da média e do desvio-padrão:

$$Q_q(x) = \mu(x) + \sigma(x) z_q, \quad (3)$$

em que $z_q = \Phi^{-1}(q)$ é o quantil de ordem q da Normal padrão $\mathcal{N}(0, 1)$ e Φ denota a sua função de distribuição acumulada.

Podemos, então, reparametrizar o modelo em função de $Q_q(x)$, escrevendo

$$\mu(x) = Q_q(x) - \sigma(x) z_q. \quad (4)$$

1.1 Verossimilhança em termos do quantil

A densidade de $Y | X = x$ é

$$f(y | x) = \frac{1}{\sqrt{2\pi} \sigma(x)} \exp\left\{-\frac{(y - \mu(x))^2}{2\sigma^2(x)}\right\}. \quad (5)$$

Substituindo $\mu(x) = Q_q(x) - \sigma(x) z_q$, obtemos

$$f(y | x) = \frac{1}{\sqrt{2\pi} \sigma(x)} \exp\left\{-\frac{(y - Q_q(x) + \sigma(x) z_q)^2}{2\sigma^2(x)}\right\}. \quad (6)$$

A log-verossimilhança para uma observação (x_i, y_i) , desprezando constantes que não dependem de Q_q , é dada por

$$\ell_i(Q_q, \sigma) = -\log \sigma(x_i) - \frac{(y_i - Q_q(x_i) + \sigma(x_i)z_q)^2}{2\sigma^2(x_i)}. \quad (7)$$

O correspondente negativo da log-verossimilhança (função de perda) é

$$\mathcal{L}_{\text{NLL}} = \sum_{i=1}^n \left[\log \sigma(x_i) + \frac{(y_i - Q_q(x_i) + \sigma(x_i)z_q)^2}{2\sigma^2(x_i)} \right]. \quad (8)$$

1.2 Caso com variância conhecida

Se assumirmos uma variância condicional conhecida e constante, $\sigma(x) \equiv \sigma_0$, o termo $\log \sigma(x_i)$ torna-se constante e pode ser ignorado na minimização. Nesse caso, a perda se reduz a

$$\mathcal{L}_{\text{NLL}} \propto \sum_{i=1}^n (y_i + \sigma_0 z_q - Q_q(x_i))^2. \quad (9)$$

Definindo a variável transformada

$$Z_i = y_i + \sigma_0 z_q, \quad (10)$$

temos

$$Z_i | X_i = x \sim \mathcal{N}(Q_q(x), \sigma_0^2), \quad (11)$$

ou seja, $Q_q(x)$ é exatamente a *média condicional* de Z_i dado $X_i = x$.

Assim, sob o modelo Normal com variância constante, podemos estimar o quantil condicional $Q_q(x)$ minimizando um *erro quadrático médio* sobre a variável transformada Z_i :

$$\mathcal{L}_{\text{MSE-quantil}} = \sum_{i=1}^n (Z_i - Q_q(x_i))^2. \quad (12)$$

1.3 Perda de quantil (pinball loss)

De forma mais geral, sem assumir Normalidade, o quantil condicional de ordem q pode ser obtido minimizando a chamada *quantile loss* ou *pinball loss*, definida para um resíduo $u = y - Q_q(x)$ por

$$\rho_q(u) = \begin{cases} q u, & \text{se } u \geq 0, \\ (q-1) u, & \text{se } u < 0. \end{cases} \quad (13)$$

A função $Q_q(\cdot)$ que minimiza

$$\mathcal{L}_{\text{pinball}} = \sum_{i=1}^n \rho_q(y_i - Q_q(x_i)) \quad (14)$$

é precisamente o quantil condicional de ordem q de $Y | X$, independentemente da distribuição de Y .

No caso particular em que o modelo Normal é bem especificado, tanto a perda $\mathcal{L}_{\text{MSE-quantil}}$ derivada da verossimilhança quanto a pinball loss $\mathcal{L}_{\text{pinball}}$ admitem o mesmo minimizador verdadeiro $Q_q(x)$. Na prática, podemos compará-las empiricamente treinando dois modelos de rede neural distintos, um com cada função de perda, e verificando se as estimativas resultantes de $Q_q(x)$ coincidem.