Universidade de São Paulo
Instituto de Matemática e Estatística
Aluna: Nara Avila Moraes
Nro. USP: 5716734

2o. semestre, 2025
Lista 03 - 21/11

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

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

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

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

$$\text{set.seed}(32)$$

$$n = 1000$$

$$x = \text{sort}(\text{runif}(n, -4, 4))$$

$$y = \frac{3}{3 + 2|x|^3} + e^{-x^2} + \cos(x)\sin(x) + 0.3\,\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0,1)$$

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

Resposta para Ítem 1.1

Encontrando a função de verossimilhança em termos do quantil condicional

Seja

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

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

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

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

$$Q_q(x) = \mu(x) + \sigma(x)\, z_q, \tag{3}$$

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

$$\mu(x) = Q_q(x) - \sigma(x)\, z_q. \tag{4}$$

0.1 Verossimilhança em termos do quantil

A densidade de $Y \mid X = x$ é

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

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

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

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

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

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

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

Descrição do experimento e configuração do modelo

Hiperparâmetros do modelo

Arquitetura

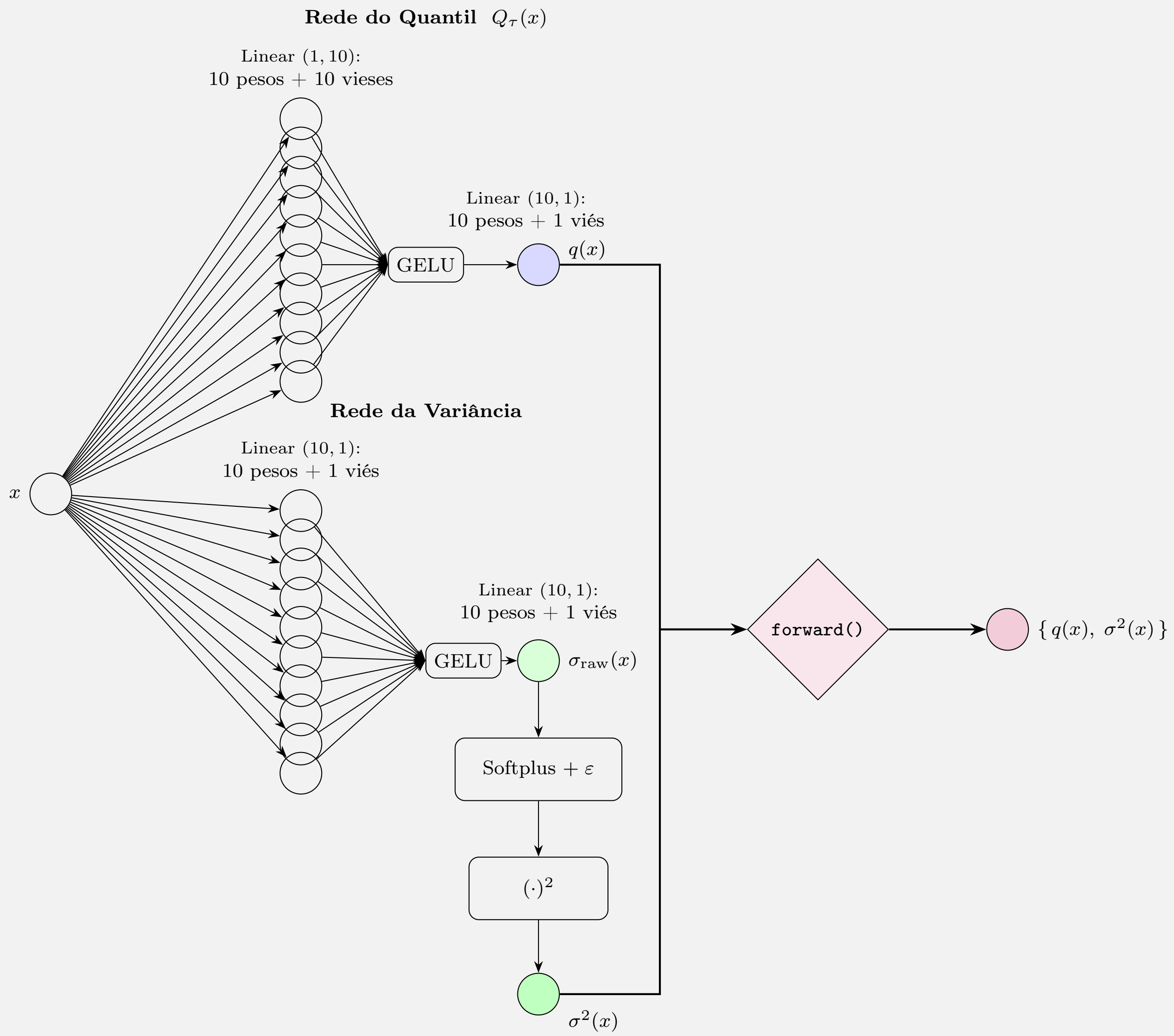

Figura 1: Arquitetura completa com 10 neurônios ocultos em cada rede. As anotações indicam os números de pesos e vieses em cada camada linear; a função `forward()` apenas combina as saídas $q(x)$ e $\sigma^2(x)$.

Tamanho do modelo

Parâmetro	**Dimensão**	**Nº de parâmetros**
q_net.0.weight	10×1	10
q_net.0.bias	10	10
q_net.2.weight	1×10	10
q_net.2.bias	1	1
sigma_net.0.weight	10×1	10
sigma_net.0.bias	10	10
sigma_net.2.weight	1×10	10
sigma_net.2.bias	1	1
Total	—	**62**

Tabela 1: Parâmetros treináveis da rede neural quantílica utilizada no experimento.

Treinamento e métricas