

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.

Ítem 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.

Ítem 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|x|^3) + e^{-x^2} + cos(x) sin(x) + 0.3 ε,    ε ~ 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

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

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

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

0.1 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

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

Descrição do experimento e configuração do modelo

Hiperparâmetros do modelo

Tabela 1: Hiperparâmetros utilizados no modelo de quantis.

Hiperparâmetro	Valor	Descrição
q	0.75	quantil alvo modelado pela rede.
num_epochs	2500	número total de épocas de treinamento.
lr	10^{-3}	taxa de aprendizado (learning rate).
input_dim	1	dimensão da entrada x .
hidden_mu	10	número de neurônios na camada oculta da rede do quantil.
hidden_sigma	10	número de neurônios na camada oculta da rede da variância.
seed	2025	semente aleatória utilizada nos experimentos.
p_train	0.7	proporção dos dados destinada ao treino.

Definição da Arquitetura

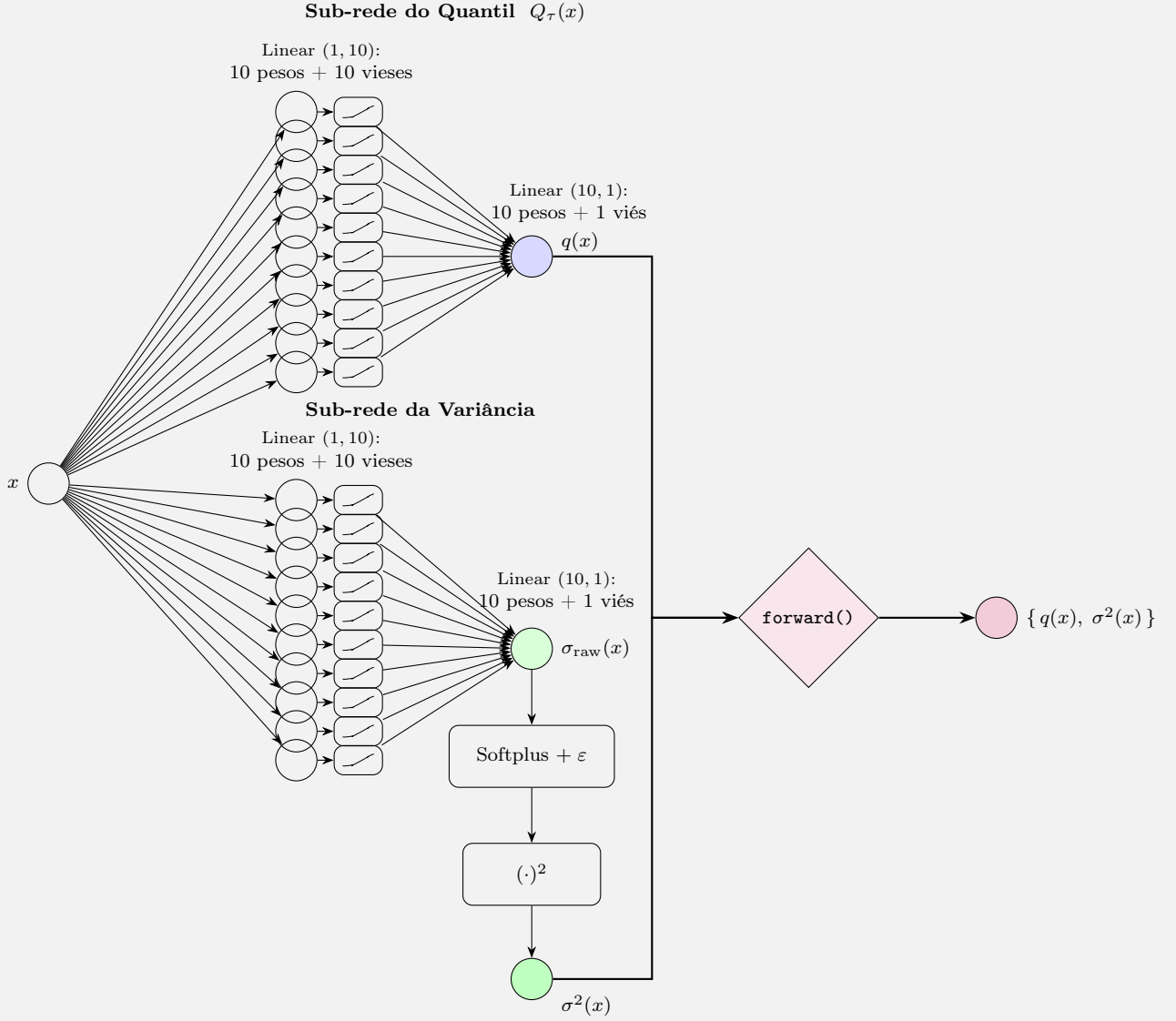


Figura 1: Arquitetura da rede neural com 10 neurônios ocultos em cada subrede.

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 2: Parâmetros treináveis da rede neural utilizada no experimento.

Treinamento e métricas

Para o treinamento foram utilizados dados sintéticos gerados a partir de uma função $f(x)$ arbitrária com ruído que segue uma variância heteroscedástica $\sigma^2(x)$, da seguinte forma:

$$Y = f(x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2(x)),$$

de modo que tanto a tendência central quanto a dispersão dos dados variam continuamente ao longo do domínio de x .

A média verdadeira utilizada foi definida por

$$f(x) = \frac{3}{3 + 2|x|^3} + e^{-x^2} + \cos(x) \sin(x).$$

A variância condicional verdadeira foi especificada por:

$$\sigma^2(x) = 0.3 + 0.8(1 + \sin(1 + 2x) + 0.2 \cos(1 + 2x))^2.$$

A heteroscedasticidade dos dados é essencial para fornecer a capacidade do modelo em aprender padrões de ruído condicional.

Os dados finais utilizados no treinamento foram obtidos por:

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2(x_i)),$$

com $n = 1000$ amostras no intervalo $[-4, 4]$, ordenadas para facilitar a visualização gráfica, e divididas em proporção 70% para treino e 30% para teste.

Treino e Resultado