Bài Tập ôn cuối kì Nhận dạng mẫu

Lưu Giang Nam

Ngày 29 tháng 12 năm 2025

1. Đo sự khác biệt giữa hai phân bố $p(x)$ (thật) và $q(x)$ (xấp xỉ):

$$KL(p||q) = \int p(x) \ln \frac{p(x)}{q(x)} \, dx \tag{1}$$

   Chứng minh các câu sau:

   (a) $KL(p||q) \neq KL(q||p)$

   (b) $KL(p||p) = 0$

   (c) $KL(p||q) \geq 0$

2. Phân phối Bernoulli:

$$\text{Bern}(x|\mu) = \mu^x (1-\mu)^{1-x} \tag{2}$$

   (a) Trung bình và phương sai của biến Bernoulli là: $\mathbb{E}[x] = \mu, \quad \text{var}[x] = \mu(1-\mu)$.

   (b) Giả sử ta có tập dữ liệu $D = \{x_1, x_2, \ldots, x_N\}$ gồm N phép thử độc lập. Khi đó, xác suất quan sát toàn bộ dữ liệu (hay còn gọi là hàm hợp lý - likelihood) được viết:

$$p(D|\mu) = \prod_{n=1}^{N} \mu^{x_n} (1-\mu)^{1-x_n} \tag{3}$$

   Sử dụng ước lượng cực đại hợp lý (MLE) chứng minh: $\mu_{\text{ML}} = \dfrac{1}{N} \displaystyle\sum_{n=1}^{N} x_n = \dfrac{m}{N}$.

3. Nếu ta thực hiện N phép thử Bernoulli độc lập, số lần thành công m tuân theo phân phối nhị thức:

$$\text{Bin}(m|N,\mu) = \binom{N}{m} \mu^m (1-\mu)^{N-m} \tag{4}$$

   Các đại lượng kỳ vọng và phương sai của m lần thành công là:

$$\mathbb{E}[m] = N\mu, \quad \text{var}[m] = N\mu(1-\mu) \tag{5}$$

4. Cho dữ liệu $D = \{x_1, \ldots, x_N\}$ độc lập, hàm hợp lý: $p(D|\mu) = \displaystyle\prod_{n=1}^{N} \prod_{k=1}^{K} \mu_k^{x_{nk}} = \prod_{k=1}^{K} \mu_k^{\sum_n x_{nk}}$ nếu định nghĩa số lần xuất hiện trạng thái k là $m_k = \sum_{n=1}^{N} x_{nk}$, $\sum_{k=1}^{K} m_k = N$. Khi đó ta có:

$$p(D|\mu) = \prod_{k=1}^{K} \mu_k^{m_k}. \tag{6}$$

   (a) Viết Log-hợp lý sử dụng Lagrange multiplier λ với ràng buộc $\sum_k \mu_k = 1$.

   (b) Chứng minh $\mu_k^{\text{ML}} = \dfrac{m_k}{N}$.

5. Giả sử ta có bộ dữ liệu $X = (x_1, x_2, \ldots, x_N)^T$, trong đó mỗi $x_n \in \mathbb{R}^D$. Mỗi quan sát x_n được giả định lấy mẫu độc lập từ phân phối Gaussian đa biến:

$$p(X|\mu,\Sigma) = \prod_{n=1}^{N} p(x_n|\mu,\Sigma), \text{ với } p(x_n|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x_n-\mu)^T \Sigma^{-1} (x_n-\mu)\right) \tag{7}$$

   Mục tiêu: tìm μ_{ML} và Σ_{ML} sao cho $p(X|\mu,\Sigma)$ là lớn nhất.

(a) Tính $\frac{\partial}{\partial \mu} \ln p(X|\mu, \Sigma)$.

(b) Tính μ_{ML} và Σ_{ML}

6. Với tập dữ liệu $(\mathbf{X}, t) = \{(\mathbf{x}_n, t_n)\}_{n=1}^{N}$, giả sử các điểm dữ liệu độc lập:

$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\big(t_n \mid \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_n), \beta^{-1}\big) \tag{8}$$

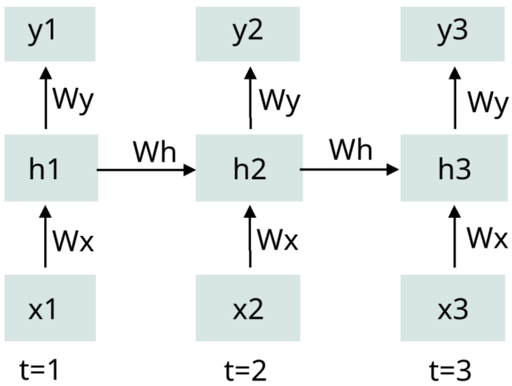
(a) Chứng minh

$$\ln p(t|w, \beta) = \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi) - \beta E_D(w) \tag{9}$$

với hàm $E_D(w) = \frac{1}{2} \sum_{n=1}^{N} \{t_n - w^T \boldsymbol{\phi}(x_n)\}^2$ chính là hàm lỗi bình phương sai số.

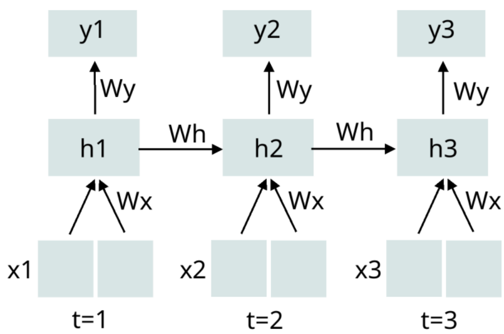
(b) Chứng minh $\mathbf{w}_{ML} = (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{\Phi}^T \mathbf{t}$.

7. Cho quá trình trải rộng (unfolding) các bước thời gian đầu tiên của một mạng RNN, trong đó cả lớp ẩn và lớp đầu ra đều là tuyến tính với $h(t) = W_x x_t + W_h h_{t-1},\ y(t) = W_y h_t$: Tính toán giá trị cho các đơn



vị ẩn h_1, h_2, h_3 và các đơn vị đầu ra y_1, y_2, y_3, với các giá trị đầu vào $x_1 = 2$, $x_2 = -0.5$, $x_3 = 1$, giả sử rằng các ma trận trọng số là $W_x = W_h = W_y = 1$.

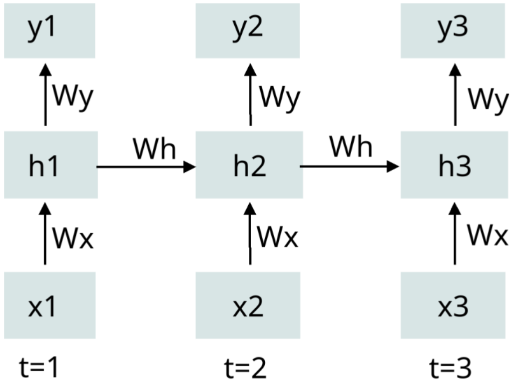
8. Cho quá trình trải rộng (unfolding) các bước thời gian đầu tiên của một mạng RNN, trong đó lớp ẩn là tuyến tính và lớp đầu ra sử dụng hàm kích hoạt sigmoid $\sigma(z) = \frac{1}{1+e^{-z}}$ như sau: $h(t) = W_x x_t + W_h h_{t-1},\ y(t) = \sigma(W_y h_t)$: Hãy tính toán giá trị của các đơn vị ẩn h_1, h_2, h_3 và các đơn vị đầu ra y_1, y_2, y_3,



với các giá trị đầu vào $x_1 = (2, -2)$, $x_2 = (0, 3.5)$, $x_3 = (1, 2.2)$, giả sử rằng các ma trận trọng số là $W_x = (1, -1)$ và $W_h = W_y = 1$.

9. Cho quá trình trải rộng (unfolding) các bước thời gian đầu tiên của một mạng RNN, trong đó lớp ẩn có hàm kích hoạt sigmoid $\sigma(z) = \frac{1}{1+e^{-z}}$ và lớp đầu ra là tuyến tính: $h(t) = \sigma(W_x x_t + W_h h_{t-1} + b_h),\ y(t) = W_y h_t + b_y$.

Hãy tính toán giá trị của các đơn vị ẩn h_1, h_2, h_3 và các đơn vị đầu ra y_1, y_2, y_3, với các giá trị đầu vào $x_1 = 18, x_2 = 9, x_3 = -8$, giả sử rằng các ma trận trọng số là $W_x = -0.1, W_h = 0.5, W_y = 0.25$ và các độ chệch (bias) là $b_h = 0.4, b_y = 0$.



10. Xét một mạng RNN đơn giản với các phương trình sau:

$$h_t = \tanh(W_x x_t + W_h h_{t-1} + b_h)$$
$$y_t = W_y h_t + b_y$$

trong đó:

- x_t là đầu vào tại thời điểm t,
- h_t là trạng thái ẩn tại thời điểm t, được thiết lập bằng 0 khi $t = 0$,
- W_h, W_x, W_y là các ma trận trọng số, và
- b_h, b_y là các độ chệch (bias).

(a) Cho các giá trị sau, hãy tính toán các trạng thái ẩn h_1 và h_2 cho chuỗi đầu vào $x_1 = 1, x_2 = 2$:

$$W_x = \begin{bmatrix} 0.6 \\ 0.1 \end{bmatrix}, \quad W_h = \begin{bmatrix} 0.5 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}, \quad b_h = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad b_y = 0.1$$

(b) Với các giá trị đã cho ở trên, số chiều (dimensionality) của h_t, y_t, và W_y trong mạng RNN này là bao nhiêu? Hãy giải thích câu trả lời của bạn.

11. Cho một tế bào LSTM với các phương trình sau:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$
$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$
$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$
$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$
$$h_t = o_t \odot \tanh(c_t)$$

Trong đó:

- $\sigma(x) = \dfrac{1}{1 + e^{-x}}$ (hàm kích hoạt sigmoid).
- $\tanh(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$ (hàm kích hoạt tanh).
- $\odot$ đại diện cho phép nhân từng phần tử (element-wise multiplication).

Hãy tính toán thủ công f_t, i_t, o_t, c_t, h_t dựa trên các giá trị sau:

- $x_t = 1, \quad h_{t-1} = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}, \quad c_{t-1} = \begin{bmatrix} 0.05 \\ 0.1 \end{bmatrix}$
- Cổng quên (Forget gate): $W_f = \begin{bmatrix} 0.4 \\ -0.2 \end{bmatrix}, U_f = \begin{bmatrix} 0.1 & 0.3 \\ 0.2 & 0.5 \end{bmatrix}, b_f = \begin{bmatrix} 0.1 \\ -0.1 \end{bmatrix}$
- Cổng vào (Input gate): $W_i = \begin{bmatrix} 0.2 \\ -0.3 \end{bmatrix}, U_i = \begin{bmatrix} 0.4 & 0.2 \\ -0.1 & 0.3 \end{bmatrix}, b_i = \begin{bmatrix} 0.05 \\ -0.05 \end{bmatrix}$

- Cổng ra (Output gate): $W_o = \begin{bmatrix} -0.3 \\ 0.5 \end{bmatrix}, U_o = \begin{bmatrix} 0.2 & -0.2 \\ 0.1 & 0.4 \end{bmatrix}, b_o = \begin{bmatrix} 0.02 \\ 0.01 \end{bmatrix}$
- Trạng thái ô nhớ ứng viên (Candidate cell state): $W_c = \begin{bmatrix} 0.3 \\ 0.1 \end{bmatrix}, U_c = \begin{bmatrix} -0.2 & 0.3 \\ 0.4 & -0.1 \end{bmatrix}, b_c = \begin{bmatrix} 0.0 \\ 0.1 \end{bmatrix}$

12. Xét một mạng nơ-ron hai lớp đơn giản. Mạng có kích thước đầu vào là 2, một lớp ẩn kích thước là 3, và kích thước đầu ra là 2. Cho đầu vào $x = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$, hai tập trọng số:

$$W_1 = \begin{bmatrix} -0.3 & 0.2 \\ 0.6 & 0.6 \\ 0.4 & 0.8 \end{bmatrix}, \quad W_2 = \begin{bmatrix} 1.8 & 0.9 & -0.5 \\ 0.5 & -1.1 & 0.2 \end{bmatrix} \tag{10}$$

và các độ chệch (biases):

$$b_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad b_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{11}$$

(a) Tính $z_1 = W_1 x + b_1$.

(b) Giả sử chúng ta sử dụng hàm Sigmoid $S(x)$ làm hàm kích hoạt cho lớp ẩn, hãy viết công thức hàm Sigmoid và tính $a_1 = S(z_1)$.

(c) Tính $z_2 = W_2 a_1 + b_2$.

(d) Sử dụng hàm Softmax $\sigma(x)$ để tính $\hat{y} = \sigma(z_2)$.

(e) Giải thích sự khác biệt giữa Sigmoid và Softmax, cũng như vai trò của chúng trong mạng nơ-ron.

13. Giả sử chúng ta có một mạng nơ-ron hai lớp được áp dụng cho bài toán phân loại nhị phân. Kiến trúc của mạng được định nghĩa như sau:

$$\begin{aligned} z_1 &= W_1 x^{(i)} + b_1 \\ a_1 &= LR(z_1) \\ z_2 &= W_2 a_1 + b_2 \\ \hat{y}^{(i)} &= \sigma(z_2) \\ L^{(i)} &= -y^{(i)} \log(\hat{y}^{(i)}) - (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \\ J &= \frac{1}{m} \sum_{i=1}^{m} L^{(i)} \end{aligned}$$

Trong đó LR là hàm Leaky ReLU và σ là hàm sigmoid:

$$LR(x) = \max(0.01x, x), \quad \sigma(x) = \frac{1}{1 + e^{-x}} \tag{12}$$

Cho tập dữ liệu $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$ với $N = 5000$. Mỗi $x^{(i)}$ có kích thước 64×1 và $y^{(i)}$ là một vô hướng. Lớp ẩn có 128 nút (z_1 có kích thước 128×1). Hãy thực hiện các yêu cầu sau:

(a) Xác định kích thước của W_1, b_1, W_2, b_2. Nếu vector hóa cho toàn bộ dữ liệu, kích thước của X và Y là bao nhiêu?

(b) Tính $\frac{\partial J}{\partial \hat{y}^{(i)}}$ và $\frac{\partial \hat{y}^{(i)}}{\partial z_2}$.

(c) Tính $\frac{\partial z_2}{\partial a_1}, \frac{\partial a_1}{\partial z_1}, \frac{\partial z_1}{\partial W_1}$.

(d) Tính $\frac{\partial J}{\partial W_1}$ (sử dụng quy tắc chuỗi và lưu ý về kích thước ma trận).

14. Giả sử chúng ta có một hàm đơn giản $f(x, y, z) = (x + y)z$. Chúng ta có thể chia nhỏ hàm này thành các phương trình $q = x + y$ và $f = qz$. Giả sử chúng ta tính toán hàm tại $x = -2, y = 5, z = -4$ và đạo hàm ngược dòng $\frac{\partial L}{\partial f} = 1$. Hãy tìm các giá trị sau theo dạng ký hiệu và dạng số:

(a) $\partial f / \partial q$

(b) $\partial q/\partial x$

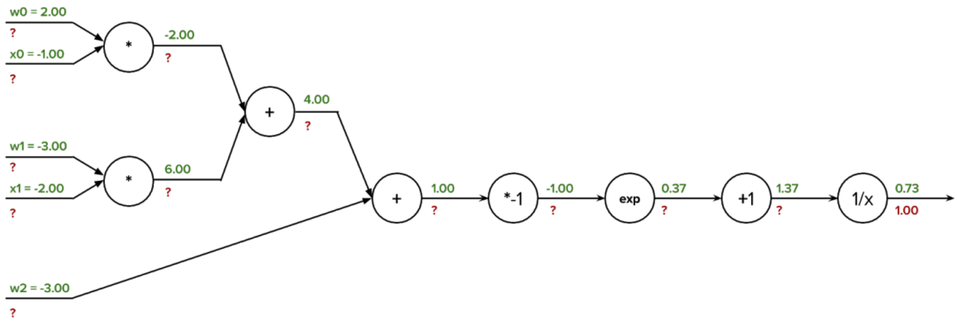
(c) $\partial q/\partial y$

(d) $\partial f/\partial z$

(e) $\partial f/\partial x$

(f) $\partial f/\partial y$

15. Bây giờ chúng ta hãy thực hiện lan truyền ngược (backpropagation) qua một nơ-ron đơn lẻ của mạng nơ-ron với hàm kích hoạt sigmoid. Cụ thể, ta định nghĩa giá trị trước kích hoạt là $z = w_0x_0 + w_1x_1 + w_2$ và giá trị kích hoạt là $a = \sigma(z) = \frac{1}{1+e^{-z}}$.



Trong đồ thị tính toán, các giá trị lan truyền xuôi đã được điền phía trên các đường kẻ, cùng với đạo hàm ngược dòng $\frac{\partial L}{\partial a}$. Hãy sử dụng thông tin này để tính toán các đạo hàm còn lại (được đánh dấu bằng dấu hỏi) trong đồ thị. Hãy tính các đạo hàm dưới dạng ký hiệu đối với các tham số đầu vào: $\dfrac{\partial a}{\partial x_0}, \dfrac{\partial a}{\partial w_0}, \dfrac{\partial a}{\partial x_1}, \dfrac{\partial a}{\partial w_1}, \dfrac{\partial a}{\partial w_2}$.

16. Giả sử chúng ta có một mạng nơ-ron hai lớp được định nghĩa như sau:

$$
\begin{aligned}
z_1 &= W_1x^{(i)} + b_1 \\
a_1 &= \text{ReLU}(z_1) \\
z_2 &= W_2a_1 + b_2 \\
\hat{y}^{(i)} &= \sigma(z_2) \\
L^{(i)} &= -\left[y^{(i)}\log(\hat{y}^{(i)}) + (1 - y^{(i)})\log(1 - \hat{y}^{(i)})\right] \\
J &= \frac{1}{m}\sum_{i=1}^{m} L^{(i)}
\end{aligned}
$$

Lưu ý rằng $x^{(i)}$ có kích thước $D_x \times 1$, $y^{(i)}$ là một vô hướng, và có m mẫu dữ liệu. Lớp ẩn có D_{a1} nút (z_1 có kích thước $D_{a1} \times 1$).

(a) Xác định kích thước của W_1, b_1, W_2, b_2. Nếu thực hiện vector hóa (vectorizing) trên nhiều mẫu dữ liệu, kích thước của X và Y sẽ là bao nhiêu? Các ma trận trọng số và độ chệch có thay đổi kích thước không?

(b) Tính $\dfrac{\partial J}{\partial \hat{y}^{(i)}}$. Kí hiệu kết quả này là $\delta_1^{(i)}$. Dựa vào đó, hãy xác định $\dfrac{\partial J}{\partial \hat{y}}$.

(c) Tính $\dfrac{\partial \hat{y}^{(i)}}{\partial z_2}$. Kí hiệu kết quả này là $\delta_2^{(i)}$.

(d) Tính $\dfrac{\partial z_2}{\partial a_1}$ là gì? Ký hiệu kết quả này là $\delta_3^{(i)}$.

(e) Tính $\dfrac{\partial a_1}{\partial z_1}$ là gì? Ký hiệu kết quả này là $\delta_4^{(i)}$.

(f) Tính $\dfrac{\partial z_1}{\partial W_1}$ là gì? Ký hiệu kết quả này là $\delta_5^{(i)}$.

(g) Tính $\dfrac{\partial J}{\partial W_1}$ là gì?

17. Xét mô hình hồi quy tuyến tính tối thiểu hóa hàm lỗi tổng bình phương có điều chuẩn (regularized sum-of-squares):

$$J(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}(\mathbf{w}^T\phi(\mathbf{x}_n) - t_n)^2 + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \tag{13}$$

trong đó $\lambda \geq 0$. Ma trận thiết kế, hàng thứ n là $\phi(\mathbf{x}_n)^T$. Vector $\mathbf{a} = (a_1, \ldots, a_N)^T$ được định nghĩa $a_n = -\frac{1}{\lambda}(\mathbf{w}^T\phi(\mathbf{x}_n) - t_n)$. Định nghĩa **Ma trận Gram** $\mathbf{K} = \Phi\Phi^T$.

(a) Đặt đạo hàm của $J(\mathbf{w})$ theo $\mathbf{w}$ bằng 0, chứng minh nghiệm $\mathbf{w}$ là tổ hợp tuyến tính của các vector $\phi(\mathbf{x}_n)$ dạng $\mathbf{w} = \Phi^T\mathbf{a}$, từ đó tính $J(\mathbf{a})$ theo $\mathbf{a}$, $\mathbf{t}$, Φ, $\mathbf{K}$.

(b) Đặt đạo hàm của $J(\mathbf{a})$ theo $\mathbf{a}$ bằng 0, tính $\mathbf{a}$ từ đó chứng minh:

$$y(\mathbf{x}) = \mathbf{w}^T\phi(\mathbf{x}) = \mathbf{k}(\mathbf{x})^T(\mathbf{K} + \lambda\mathbf{I}_N)^{-1}\mathbf{t} \tag{14}$$

18. Kiểm chứng tính hợp lệ của các Kernel sau, tìm ánh xạ đặc trưng tương ứng.

(a) Xét hàm kernel: $k(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z})^2$.

(b) Xét kernel tổng quát hơn với hằng số $c > 0$:

$$k(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z} + c)^2 \tag{15}$$

(c) Giả sử ta có hai kernel hợp lệ:

- Linear Kernel: $k_1(\mathbf{x}, \mathbf{z}) = \mathbf{x}^T\mathbf{z}$ (đặc trưng tuyến tính).
- Quadratic Kernel: $k_2(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z})^2$ (đặc trưng bậc 2).

Kernel mới $k(\mathbf{x}, \mathbf{z}) = k_1(\mathbf{x}, \mathbf{z}) + k_2(\mathbf{x}, \mathbf{z})$ có hợp lệ không?

(d) Gaussian Kernel 1 chiều (giản lược):

$$k(x, z) = \exp(-(x - z)^2) \tag{16}$$

(e) Cho A_1, A_2 là các tập hợp con, kernel cho dữ liệu ký hiệu

$$k(A_1, A_2) = 2^{|A_1 \cap A_2|} \tag{17}$$