

Table 1: Dataset

The dataset consists of five samples x_1, x_2, x_3, x_4, x_5 . For each sample, we can observe the features A_1, A_2, A_3 and the corresponding response.

Solution:

$$\text{记 } S = \{x_1, x_2, x_3, x_4, x_5\}$$

$$\begin{aligned} \text{Entropy}(S) &= -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) \\ &= 0.971 \end{aligned}$$

$$\text{Gain}(S, A_1) = \text{Entropy}(S) - \sum_{\text{value}(A_1)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$\begin{aligned} &= \text{Entropy}(S) - \frac{4}{5} \text{Entropy}(A_1^+) - \frac{1}{5} \text{Entropy}(A_1^-) \\ &= 0.971 - \frac{4}{5} \times 1 - \frac{1}{5} \times 0 \end{aligned}$$

$$= 0.171$$

$$\text{Gain}(S, A_2) = \text{Entropy}(S) - \sum_{\text{value}(A_2)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0.971 - \frac{3}{5} \times \text{Entropy}(A_2^+) - \frac{2}{5} \text{Entropy}(A_2^-)$$

$$= 0.971 - \frac{3}{5} \times 0.918 - \frac{2}{5} \times 0$$

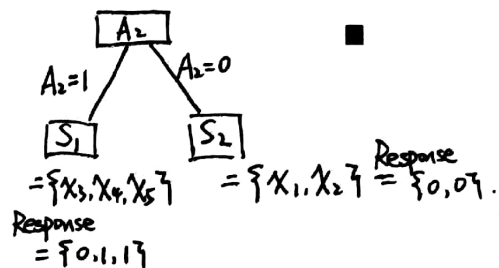
$$= 0.420$$

$$\text{Gain}(S, A_3) = \text{Entropy}(S) - \sum_{\text{value}(A_3)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0.971 - \frac{2}{5} \times \text{Entropy}(A_3^+) - \frac{3}{5} \text{Entropy}(A_3^-)$$

$$= 0.971 - \frac{2}{5} \times 1 - \frac{3}{5} \times 0.918$$

$$= 0.020 \quad \text{故选择 } A_2.$$



$$\text{对 } S_1 = \{x_3, x_4, x_5\}$$

$$\begin{aligned} \text{Entropy}(S_1) &= -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} \\ &= 0.918 \end{aligned}$$

$$\text{Gain}(S_1, A_1) = \text{Entropy}(S_1) - \sum_{\text{value}(A_1)} \frac{|S_v|}{|S_1|} \text{Entropy}(S_v)$$

$$= 0.918 - \frac{1}{3} \times 0 - \frac{2}{3} \times 0$$

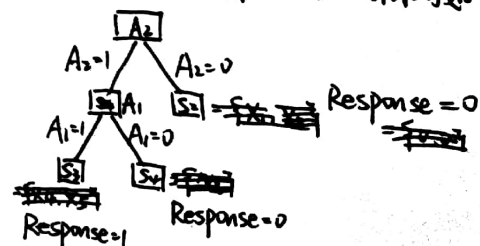
$$= 0.918$$

$$\text{Gain}(S_1, A_3) = \text{Entropy}(S_1) - \sum_{\text{value}(A_3)} \frac{|S_v|}{|S_1|} \text{Entropy}(S_v)$$

$$= 0.918 - \frac{1}{3} \times 0 - \frac{2}{3} \times 1$$

$$= 0.251$$

故选择 A_1 ，最终的决策树如下：



$$g(x) = - \sum_{i=1}^n H_i \log(f_i(x)),$$

where $H \in \mathbb{R}^n$ is a one-hot vector.

Solution:

$$\begin{aligned} 1. \quad \frac{\partial f_i(x)}{\partial x_i} &= \frac{\exp(x_i) \sum_{k=1}^n \exp(x_k) - (\exp(x_i))^2}{(\sum_{k=1}^n \exp(x_k))^2} \\ &= \frac{\exp(x_i)}{\sum_{k=1}^n \exp(x_k)} - \left(\frac{\exp(x_i)}{\sum_{k=1}^n \exp(x_k)} \right)^2 \\ &= f_i(x) - (f_i(x))^2 \end{aligned}$$

$$\begin{aligned} \frac{\partial f_i(x)}{\partial x_j} &= \frac{-\exp(x_i) \exp(x_j)}{(\sum_{k=1}^n \exp(x_k))^2} \\ &= -f_i(x) f_j(x) \end{aligned}$$

$$\begin{aligned} \nabla f(x) &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} f_1 - f_1^2 & -f_1 f_2 & \dots & -f_1 f_n \\ -f_1 f_2 & f_2 - f_2^2 & \dots & -f_2 f_n \\ \vdots & \vdots & \ddots & \vdots \\ -f_1 f_n & -f_2 f_n & \dots & f_n - f_n^2 \end{bmatrix} \end{aligned}$$

$$Df(x) = (\nabla f(x))^T = \nabla f(x)$$

$$\begin{aligned} 2. \quad f_i(x-c) &= \frac{\exp(x_i - c)}{\sum_{k=1}^n \exp(x_k - c)} \\ &= \frac{\exp(x_i) / \exp(c)}{\sum_{k=1}^n \exp(x_k) / \exp(c)} \\ &= \frac{\exp(x_i)}{\sum_{k=1}^n \exp(x_k)} = f_i(x) \end{aligned}$$

$$\therefore f(x) = f(x-c).$$

$$x_i - c = x_i - \max_k \{x_k\} \leq 0$$

在进行运算时(求导等)需要控制指数的大小, 防止(如 f_i^2 或 $f_i f_j$) 溢出.

$\exp(x_i - c) \leq 1$ 则可以控制大小.

$$\begin{aligned} 3. \quad \frac{\partial g(x)}{\partial x_j} &= - \sum_{i=1}^n H_i \frac{\partial \log(f_i(x))}{\partial x_j} \\ &= - \sum_{i=1}^n H_i \frac{\partial \log(f_i(x))}{\partial f_i(x)} \cdot \frac{\partial f_i(x)}{\partial x_j} \\ &= - \sum_{i=1}^n H_i \frac{1}{f_i(x)} \cdot \frac{\partial f_i(x)}{\partial x_j} \\ &= - \sum_{i=1, i \neq j}^n H_i \frac{-f_i(x) f_j(x)}{f_i(x)} - H_j \frac{f_i(x) - f_j^2(x)}{f_j(x)} \\ &= \sum_{i=1, i \neq j}^n H_i f_j(x) + H_j f_j(x) - H_j \\ &= f_j(x) - H_j \end{aligned}$$



扫描全能王 创建

- (a) The convolutional layer parameters are denoted as "conv(filter size)-(number of filters)".
- (b) The fully connected layer parameters are denoted as "FC(number of neurons)".
- (c) The window size of pooling layers is 3.
- (d) The stride of convolutional layers is 1.
- (e) The stride of pooling layers is 3.
- (f) There is no padding in both convolutional and pooling layers.
- (g) For convenience, we assume that there is no activation function and bias.

Suppose that the input is a **386 × 386 RGB** image. Please derive the size of all feature maps and the number of parameters.

conv3-64	max pool	conv3-256	conv1-512	max pool	FC-2048	FC-1000
----------	----------	-----------	-----------	----------	---------	---------

Table 2: The architecture of convolutional neural network

Solution:

$$1. \frac{\partial f_i(x)}{\partial x_i} = \frac{1}{n}.$$

$$\nabla f_i(x) = \left(\frac{1}{n}, \dots, \frac{1}{n} \right)^T.$$



3. 2.

con 1a) 先考虑二维 $\{x_1, x_2\}$ 的情况

$$f_2(x) = \max\{x_1, x_2\} \\ = \frac{1}{2}(x_1 + x_2 + |x_1 - x_2|)$$

在 $x_1 = x_2$ 处 $|x_1 - x_2|$ 不可微

$x_1 \neq x_2$ 处 $x_1 + x_2, |x_1 - x_2|$ 均可微,

故 $f_2(x)$ 可微.

考虑多维情况

设 x_i, x_j 为 x_1, x_2, \dots, x_n 中最大的两个数

$$f_2(x) = \max\{x_1, x_2, \dots, x_n\} = \max\{x_i, x_j\} \\ = \frac{1}{2}(x_i + x_j + |x_i - x_j|)$$

同上讨论知在 $x_i = x_j$ 时 $f_2(x)$ 不可微

$x_i \neq x_j$ 处 $f_2(x)$ 可微.

故可微点组合的集合 C 为

$$C = \{(x_1, x_2, \dots, x_n) \mid x_j = \max_i \{x_i\}, x_k = \max_{i, i \neq j} \{x_i\}, x_j \neq x_k\}$$

(b)

$$\text{令 } x_j = \max_i \{x_i\}$$

选取 $d(x)$ 为

$$d(x)_{ij} = \begin{cases} 0, & \text{其他} \\ 1, & i=j \end{cases}$$

则

$$\cancel{f_2(x)} f_2(x) + \langle d(x), y - x \rangle$$

$$= f_2(x) + y_j - x_j$$

$$= x_j + y_j - x_j = y_j \leq \max_i \{y_i\} = f_2(y).$$



3.

	feature maps	parameters
Conv3-64	$64 \times (386-2) \times (386-2)$ $= 9437184$	$3 \times 9 \times 64$ $= 1728$

max-pool	$64 \times \frac{384}{3} \times \frac{384}{3}$ $= 1048576$	无
----------	---------------------------------------------------------------	---

Conv3-256	$256 \times (\frac{384-2}{3}) \times (\frac{384-2}{3})$ $= 4064256$	$64 \times 9 \times 256$ $= 147456$
-----------	------------------------------------------------------------------------	----------------------------------------

Conv1-512	$512 \times (\frac{126-2}{3}) \times (\frac{126-2}{3})$ $= 8128512$	$256 \times 1 \times 512$ $= 131072$
-----------	------------------------------------------------------------------------	-----------------------------------------

max-pool	$512 \times \frac{126}{3} \times \frac{126}{3}$ $= 903168$	无
----------	---------------------------------------------------------------	---

FC-2048	2048	$903168 \times 2048 \times 3$ $= 1849688064$
---------	------	-------------------------------------------------

FC-1000	1000	2048×1000 $= 2048000$
---------	------	-----------------------------------

经过第一个FC后的特征已完全不区分RGB三个通道，故无需设计x3的权重参数

王



Exercise 4: Matrix Calculus 20pts

Let $L = f(h(Ax + b))$, where $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $f: \mathbb{R}^m \rightarrow \mathbb{R}$. Define $z = Ax + b \in \mathbb{R}^m$ and $w = h(z) = (\sigma(z_1), \dots, \sigma(z_m))^T$, where z_i is the i^{th} component of z and

$$\sigma(z_i) = \frac{1}{1 + \exp(-z_i)}.$$

Assume $\nabla_w f$ is known.

1. Please derive $\nabla_x L$.
2. Please derive

$$\nabla_A L = \begin{bmatrix} \frac{\partial L}{\partial A_{11}} & \cdots & \frac{\partial L}{\partial A_{1j}} & \cdots & \frac{\partial L}{\partial A_{1n}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial L}{\partial A_{i1}} & \cdots & \frac{\partial L}{\partial A_{ij}} & \cdots & \frac{\partial L}{\partial A_{in}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial L}{\partial A_{m1}} & \cdots & \frac{\partial L}{\partial A_{mj}} & \cdots & \frac{\partial L}{\partial A_{mn}} \end{bmatrix},$$

where $A_{i,j}$ is the entry in the i^{th} row, j^{th} column of the matrix A .

Solution:

1.

$$\nabla_x L = \nabla_x z^T \nabla_z w^T \nabla_w f$$

$$\nabla_x z = A^T$$

$$(\nabla_z w)_{ij} = \begin{cases} \frac{\exp(-z_j)}{(1 + \exp(-z_j))^2}, & i=j \\ 0, & i \neq j \end{cases}$$

$$\nabla_x L = A^T \begin{pmatrix} \frac{\exp(-z_1)}{(1 + \exp(-z_1))^2} & \cdots & \frac{\exp(-z_m)}{(1 + \exp(-z_m))^2} \end{pmatrix} \nabla_w f$$

2.

$$\begin{aligned} \frac{\partial L}{\partial A_{ij}} &= \frac{\partial f}{\partial w_i} \cdot \frac{\partial w_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial A_{ij}} \\ &= (\nabla_w f)_i \frac{\exp(-z_i)}{(1 + \exp(-z_i))^2} x_j \end{aligned}$$

PBI7000215
张博文



扫描全能王 创建