結構化機器學習模型及其應用

第二次報告

系所:應數所大數據組
學生:張格恩
學號:7106053112

目錄

一、Representation

c_{11}	c_{12}	c_{21}	c_{22}	a_1	b_1	a_2	b_2
E	2	A	C	1	0	6	2
5	B	6	3	1	0	6	2

……

a_{15}	b_{15}	a_{16}	b_{16}
4	4	E	E
4	4	E	E

Feature(x) :

$x = [a_1, b_1, \ldots, a_{10}, b_{10}, a_{13}, b_{13}, \ldots, a_{16}, b_{16}]$ (28 Dim) ， a_i , b_i =0~E $\forall i$

$x \longrightarrow [x_1, x_2, \ldots, x_{111}, x_{112}]$ (112 Dim) ， $x_i = 0 \; or \; 1 \quad \forall i$

Label($y = [y_1, y_2]$) :

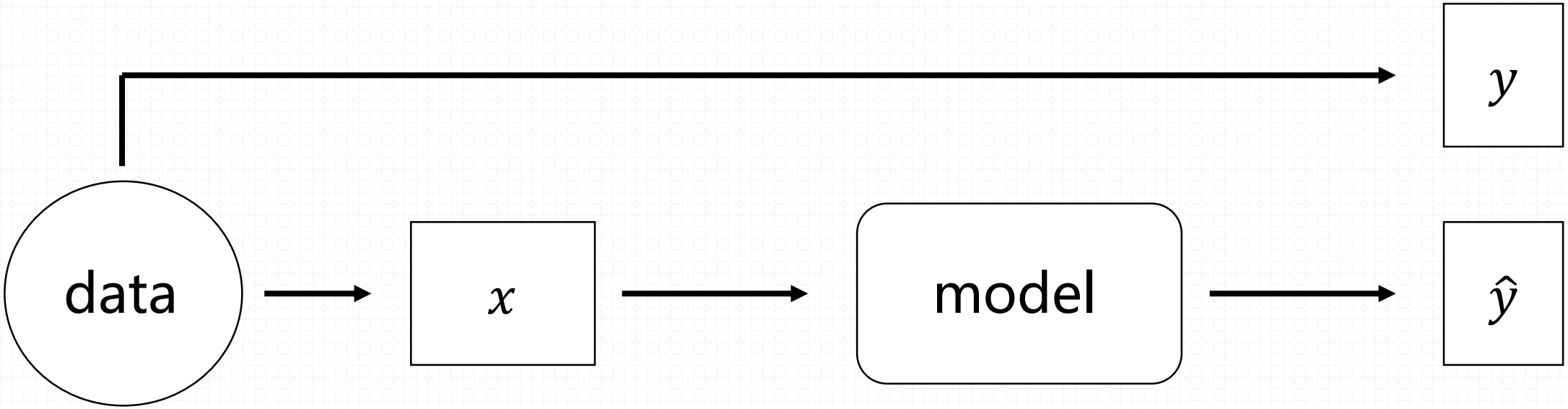
$y = [y_1, y_2] = [c_{11}, c_{12}, c_{21}, c_{22}] \longrightarrow y = 16$ bit$(000\cdots 000 \sim 111\cdots 111)$

$y \longrightarrow (y_0, y_1, \ldots, y_{14}, y_{15})$ $y_i = 0$ or 1 $\forall i$=0~15

一、Representation

Objective:

$$\min \sum_i sign(\hat{y}_i, y_i) \quad , sign(\hat{y}_i, y_i) = \begin{cases} 0, & \hat{y}_i = y_i \\ 1, & \hat{y}_i \neq y_i \end{cases} \qquad i: \text{樣本數量}$$



一、Representation

Feature(x) :

$x \longrightarrow [x_1, x_2, \ldots, x_{111}, x_{112}]$ (112 Dim) ， $x_i = 0 \; or \; 1 \quad \forall i$

Reason:

(1)增加bit of feature的相關性

(2)降低feature distribution的variance

Label($y = [y_1, y_2]$) :

$y \longrightarrow (\, y_0 \,, y_1, \ldots, y_{14}, y_{15} \,)$ $y_i = 0$ or 1 $\forall i$=0~15

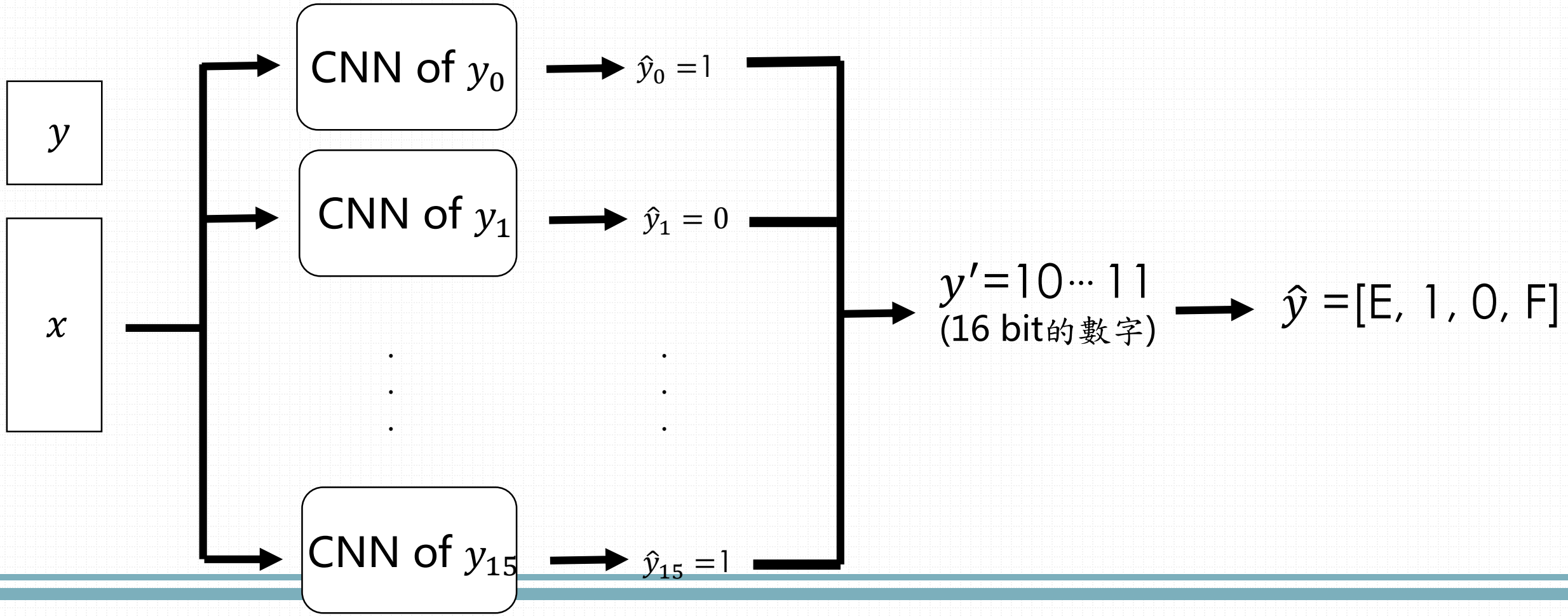
assume y_i are less relevant to each i，let problem(classification:65536) simplify 16 problem(classification:0/1)

Reason:

(1)降低問題複雜度

二、Model

Convolutional Neural Network(CNN) for y

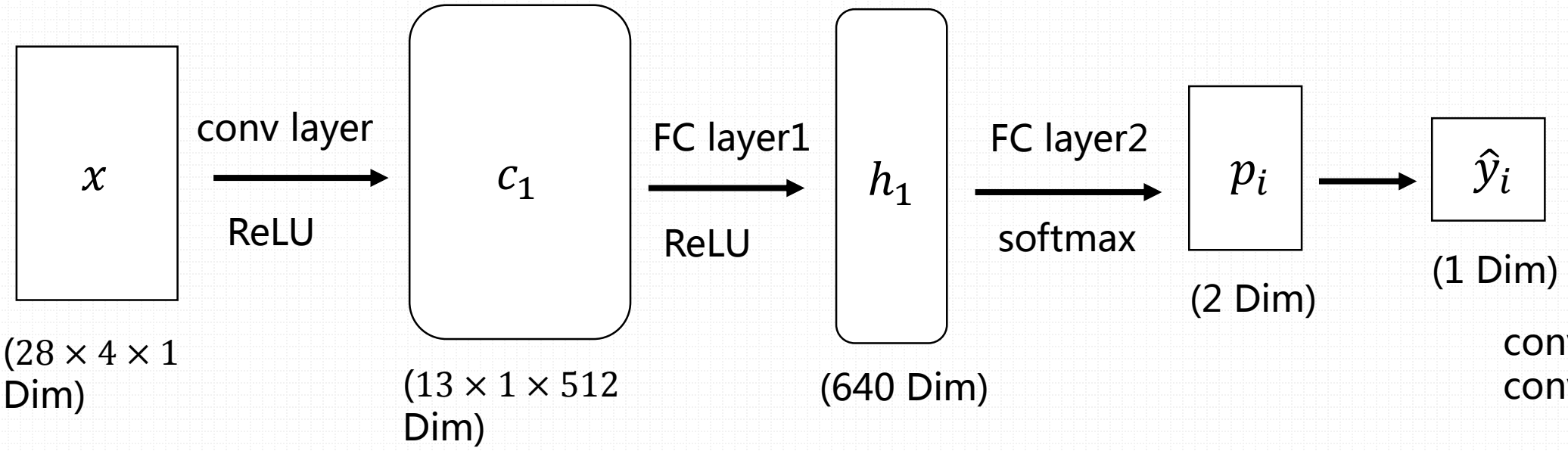


二、Model

Convolutional Neural Network(CNN) for each y_i

feature(x) : $x = [(x_{1,1}, x_{1,2}, x_{1,3}, x_{1,4}), \ldots, (x_{14,1}, x_{14,2}, x_{14,3}, x_{14,4})]$ $(28 \times 4$ Dim$)$

label(y) : $y_i = 0$ or 1 $for\ each\ i$=0~15



conv layer:
convolutional layer

FC layer:
fully connected layer

二、Model

Reasons of Convolutional Neural Network(CNN)

(1)將feature中的雜訊去除
(2)進行特徵選擇(Feature selection)，強化重要feature的權重

Parameters of Convolutional Neural Network(CNN)

(1)filter: size-4 × 4，numbers-512
希望能在卷積運算時，能將原始訊息的形狀考慮進來
(2)no dropout、Adam optimizer

二、Model

Loss function of Convolutional Neural Network(CNN) for each y_i

Y_i : the true label for y_i (i bit)
$\hat{Y}_i$: the output of NN model for y_i (i bit)
C : the matrix, values are 1

loss: $L(W) = -Tr(Y_i^T \log(\hat{Y}_i)) - Tr((C - Y_i)^T \log(C - \hat{Y}_i))$

THE END

感謝聆聽