Homework3 Report Template

Professor Pei-Yuan Wu
EE5184 - Machine Learning

系級: 資料科學碩一　　姓名：陳庭安　　學號：R07946007

1. (1%) 請說明你實作的 CNN model，其模型架構、訓練過程和準確率為何？
   (a) 模型架構

(Convolution + Max Pooling + Batch Normalization + Dropout) * 5 層

Convolution Layers

	Channels	Filters	Strides	Activation function
1	64	(3, 3)	(1, 1)	Relu
2	64	(2, 2)	(1, 1)	Relu
3	128	(2, 2)	(1, 1)	Relu
4	256	(2, 2)	(1, 1)	Relu
5	256	(2, 2)	(1, 1)	Relu

Max Pooling
    Pool size: (2, 2)
Batch Normalization
Dropout
    Dropout rate: 0.2

(FC + Batch Normalization + Dropout) * 4 層

Fully connected layers
    Nodes: 256, 128, 64, 32
Batch Normalization
Dropout
    Dropout rate = 0.2
Activation function: Relu

FC * 1 層

Nodes: 7

Total parameters: 1,110,919

(b) 訓練過程

Epochs = 50, Batch size = 50, Optimizer = Adagrad, Dropout rate (如上)

(c) 準確率

Training	Validating	Testing (Public)	Testing (Private)
0.5333	0.6149	0.60657	0.62050

2. (1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model，其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼？

(a) 模型架構

FC * 1 層

Nodes: 430

Activation function: Relu

(FC + Batch Normalization + Dropout) * 6 層

Nodes: 128, 128, 64, 64, 32, 32

Dropout rate: 0.2

Activation function: Relu

FC * 1 層

Nodes: 7

Activation function: Softmax

Total parameters: 1,080,405

(b) 訓練過程

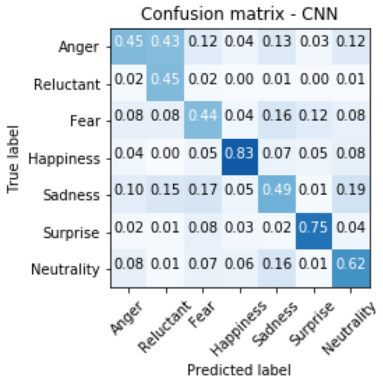
Epochs = 50, Batch size = 50, Optimizer = Adagrad, Dropout rate (如上)

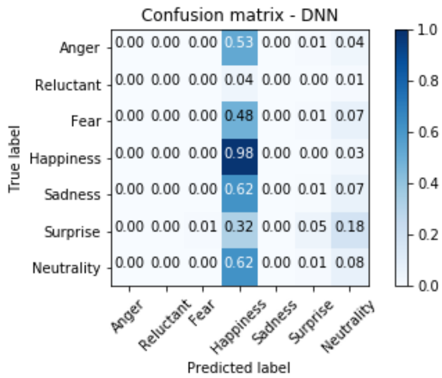
(c) 準確率

Training	Validating	Testing (Public)	Testing (Private)
0.3094	0.2654	0.26330	0.26469

(d) 在參數接近、相同超參數值的情況下，CNN 在圖形識別上，遠勝 DNN。原因是 DNN 是將圖片 Pixels 展成向量做為 inputs，但這就忽視了圖片相鄰 pixels 的關聯性；然而 CNN 便利用了 Filters 滑動，去捕捉到圖片當中的部分 pattern，因此，CNN 的 performance 高過 DNN 許多的結果並不令人意外。

3. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？ 並說明你觀察到了什麼？ [繪出 confusion matrix 分析]





由左圖 CNN 預測結果知，高興的表情最能被辨識出，而生氣、厭惡、恐懼、難過較容易被誤判。生氣的表情特別容易被辨識成厭惡。而由右圖 DNN 的預測結果可知，雖然高興的表情的識別準確率極高，但 DNN 訓練出來的模型是幾乎將所有表情都辨識成高興了，這就驗證了前一題所提及的，DNN 無法成功捕捉到圖片當中的 pattern。

ML HW3.

4.

$$[(k_i \times k_i) \times c_i + 1] \times c_{i-1}$$

(a) Layer A : $\left[(2\times 2)\times 5 + 1\right] \times 6 = 126$

filter_size #in_channel #bias # out_channel

Layer B: $\left[(2\times 2)\times 6 + 1\right] \times 4 = 100$

(b) Layer A:

$(\times)$: $\left\{(2\times 2) \times \left(\frac{8-2}{3}+1\right)^2\right\} \times 5 \times 6 = (4\times 9) \times 5 \times 6 = 1080.$

$(+)$: $\left\{(2\times 2\times 5 - 1) \times 9\right\} \times 6 = 1026.$

Layer B:

$(\times)$: $\left\{(2\times 2) \times \left(\left\lfloor\frac{3-2}{2}\right\rfloor + 1\right)^2\right\} \times 6 \times 4 = (4 \times (0+1)^2) \times 24$

$= 96$

$(+)$: $\left\{(2\times 2\times 6 - 1)\times 1\right\} \times 4 = 92$

(c)

$$O\left(\sum_{i=1}^{L} \left\{(k_i)^2 \times \left(\left\lfloor \frac{n_i + 2p_i - k_i}{s_i}\right\rfloor + 1\right)^2\right\} \times c_i \times c_{i-1}\right.$$

$$\left. + \left\{(k_i^2 \times c_i - 1) \times \left(\left\lfloor\frac{n_i + 2p_i - k_i}{s_i}\right\rfloor + 1\right)^2\right\} \times c_{i-1}\right)$$

where c_0 is the channel size of input

5.

(a) $\hat{\Sigma}\ X = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 8 & 5 \\ \vdots & \vdots & \vdots \\ 10 & 11 & 7 \end{bmatrix}_{10\times 3} = [\underline{c}_1\ \underline{c}_2\ \underline{c}_3].$

$X' = [\underline{c}_1 - \text{mean}(\underline{c}_1)\ \ \underline{c}_2 - \text{mean}(\underline{c}_2)\ \ \underline{c}_3 - \text{mean}(\underline{c}_3)]$. — centralize

$$\Sigma_{X'} = \text{cov}(X') \doteq \begin{bmatrix} 13.3778 & 0.5556 & 3.6444 \\ 0.5556 & 13.5556 & 3.2222 \\ 3.6444 & 3.2222 & 9.0667 \end{bmatrix},$$

where the degree of freedom = 1,

i.e. $\text{var}(\underline{x}_1) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{1i} - \bar{x}_{1\cdot})^2$ rather than $\frac{1}{n}\sum_{i=1}^{n}(x_{1i} - \bar{x}_{1\cdot})^2$ here

$\det(\Sigma_x - \lambda I) = 0 \Rightarrow \lambda \doteq 17.00,\ 12.92,\ 6.08$

$\Rightarrow$ Eigenvectors of Σx, i.e., pricinple axes Q

$$\doteq \begin{bmatrix} -0.6166 & -0.6782 & 0.3999 \\ -0.5888 & 0.7344 & 0.3376 \\ -0.5226 & -0.0273 & -0.8521 \end{bmatrix}.$$

Eigenvectors of Σx

Eigenvalues are 17.00, 12.92, 6.08 respectively.

5.

ans. of (a)
$\uparrow$

(b) PC Matrix / Score Matrix = Data Matrix (Scaled) × Priciple axes / Loading matrix.

$$Z = X' Q$$

$$\doteqdot \begin{bmatrix} 7.1866 & -1.3732 & -2.2510 \\ 0.7587 & 0.9440 & -0.7302 \\ -3.0703 & 4.45059 & 3.1883 \\ 2.6085 & 2.97853 & -1.9298 \\ -1.8230 & 4.75401 & 4.2516 \\ 3.3546 & -3.9190 & 2.5276 \\ -4.4164 & -2.55601 & -2.1395 \\ 3.4657 & 1.7132 & 2.2785 \\ -2.3136 & -6.03372 & 0.2039 \\ -5.7525 & -0.9765 & 0.9774 \end{bmatrix}$$

$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ PC1 & PC2 & PC3 \end{array}$$

(c). Reconstruction Error in L2-norm

$$\| \overbrace{[-4.4\ \ -6\ \ -1.8]}^{X'} - \overbrace{[7.1866\ \ -1.3732\ \ \ 0]}^{PC1\ \&\ PC2\ \text{in}\ Z} \|_2$$

$$+ \| [-1.4\ \ 0\ \ 0.2] - [0.7587\ \ 0.9440\ \ 0] \|_2$$

$$\vdots$$

$$+ \| [4.6\ \ 3.0\ \ 2.2] - [-5.7525\ \ -0.9765\ \ 0] \|_2$$

$$\doteqdot 12.61 + 2.36 + 4.28 + 7.62 + 3.38 + 4.19 + 9.40$$

$$+ 7.20 + 8.56 + 11.31$$

$$\doteqdot 70.9$$