Homework3 Report

Professor Pei-Yuan Wu
EE5184 - Machine Learning

姓名：楊宗賢
學號：B06901031

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?

我訓練了三個CNN model，命名為model_3、model_5、model_6，細節如下:

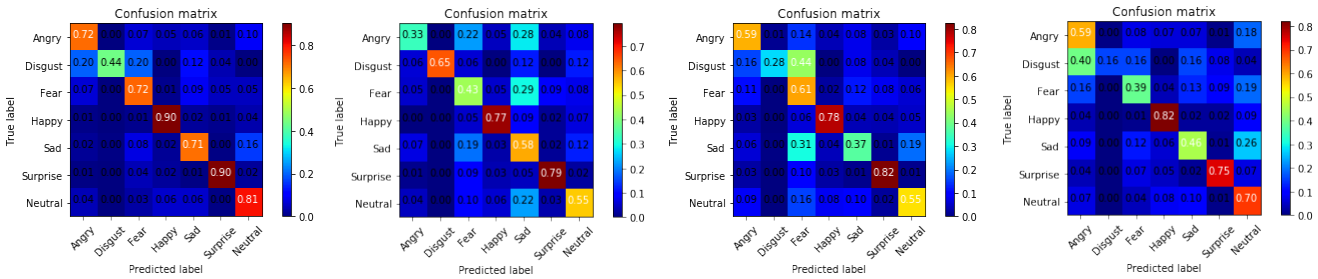
model_3	model_5	model_6
Conv, 3x3, 120	Conv, 3x3, 160	Conv, 3x3, 180
Conv, 3x3, 120	Conv, 3x3, 160	Conv, 3x3, 180
Maxpooling, 2	Maxpooling, 2	Maxpooling, 2
Dropout: 0.1	Dropout: 0.2	Dropout: 0.2
Conv, 2x2, 240	Conv, 2x2, 320	Conv, 2x2, 360
Conv, 2x2, 240	Conv, 2x2, 320	Conv, 2x2, 360
Maxpooling, 2	Maxpooling, 2	Maxpooling, 2
Dropout: 0.1	Dropout: 0.2	Dropout: 0.2
Conv, 3x3, 480	Conv, 2x2, 480	Conv, 2x2, 540
Conv, 3x3, 480	Conv, 2x2, 480	Conv, 2x2, 540
Maxpooling, 2	Conv, 2x2, 480	Conv, 2x2, 540
Dropout: 0.1	Maxpooling, 2	Maxpooling, 2
Dense: 720	Dropout: 0.3	Dropout: 0.2
Dropout: 0.3	Conv, 2x2, 640	Conv, 2x2, 720
Dense: 720	Conv, 2x2, 640	Conv, 2x2, 720
Dropout: 0.3	Conv, 2x2, 640	Conv, 2x2, 720
Dense: 7	Maxpooling, 2	Maxpooling, 2
	Dropout: 0.3	Dropout: 0.3
	Dense: 1280	Dense: 2560
	Dropout: 0.3	Dropout: 0.3
	Dense: 960	Dense: 960
	Dropout: 0.3	Dropout: 0.3
	Dense: 7	Dense: 7
(Conv padding:none)	(Conv padding:same)	(Conv padding:same)
訓練方式		
80 epochs	65 epochs	100 epochs
batch size: 100	batch size: 150	batch size: 144
準確率		
training: 0.9607	training: 0.9386	training: 0.9639
validation: 0.5800	validation: 0.6169	validation: 0.6263
Kaggle private: 0.60490	Kaggle private: 0.61493	Kaggle private: 0.62412

三個 model 以相同的 weight 合取 ensemble 後，
validation accuracy=0.7950，Kaggle private score=0.67038。

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？

我修改了 model_6，作成了約有 2900 萬個參數的 DNN model：
模型架構：[input=2304]→880→1760→2640→3520→2560→960→7
訓練過程：同 model_6，100 epochs，batch size=144
準確率：training=0.3959 validation=0.3404 Kaggle private=0.32738
此 DNN 的 Confusion Matrix 顯示，Surprise 認出 0.76，Happy 認出 0.53，其餘的 class 都認不出來，連 training data 都無法 fitting。上題中，training data 至少都有 0.9 以上，顯見卷積層的有無對圖像辨識的成敗有多重要。

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？



最左為 ensembled model，其餘依序為三個 model 各自的 Confusion Matrix。

Disgust 的 class 最容易與其他 class 混淆，各有 0.2 的 Disgust 被分成了 Angry 和 Fear。另外 Fear 被認成 Sad、Sad 或 Angry 被認成 Neutral 的情形也很顯著，反倒是 Happy 和 Surprise 不太會被認錯。

按照一般社會常識來推論，正面情緒的表情會有很顯著的特徵，而負面的情緒(如 Angry、Fear、Sad 等)可能都同樣表現出一個「激動」的表情，自然會互相混淆，只能到 0.7 的 Accuracy。

training set 中的 Disgust class 佔比實在太小了，層數高的 model_5 與 model_6 都幾乎完全認不出來，因此我將層數雖少卻能將 Disgust 認到 0.6 的 model_3 加入 ensemble，平衡了另兩個 model 在 Disgust class 上的重大缺點。

4. (1.5%,each 0.5%)CNN time/space complexity:

4.

a.

Layer A: $\underbrace{(2\times2\times5}_{\text{weight}}+\underbrace{1)}_{\text{bias}}\times 6 = 126$ (parameters)

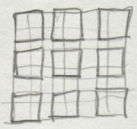
Layer B: $\underbrace{(2\times2\times6}_{\text{weight}}+\underbrace{1)}_{\text{bias}}\times 4 = 100$ (parameters)

b.

Layer A:

each kernel-size calculations takes $\sum_{i=1}^{2\times2\times5} w_i \cdot x_i$,

making 19 additions、 20 multiplications;



there are 6 filters, each filter perform 3×3 times,

so $19\times6\times3\times3 = 1026$ additions and

$20\times6\times3\times3 = 1080$ multiplications are needed.

Layer B:

there are 4 filters, each kernel has $2\times2\times6=24$ weight params,

each filter perform 1×1 time 

so $(24-1)\times4\times1\times1 = 92$ additions and

$24\times4\times1\times1 = 96$ multiplications are needed.

c.

each layer: $O\left([k^2C_{i-1}+(k^2C_{i-1}-1)]\times C_i\times\left[L\frac{n+2p-k}{s}\right]+1\right]\right)$

$= O\left((k^2(n+2p-k)\div s)\, C_{i-1}{}^2\, C_i\right)$

denote $k^2(n+2p-k)\div s$ as K_i for layer i

$\Rightarrow$ total time complexity $= O\left(\sum_{i=2}^{d} K_i (C_{i-1})^2 C_i\right)$

5. (1.5%,each 0.5%)PCA practice:

5.

a. $\mu = \begin{bmatrix} 54 \\ 80 \\ 48 \end{bmatrix} \div 10 = \begin{bmatrix} 5.4 \\ 8.0 \\ 4.8 \end{bmatrix}$

$$\Rightarrow N\Sigma = \begin{bmatrix} 19.36 & 26.4 & 7.92 \\ 26.4 & 36 & 10.8 \\ 7.92 & 10.8 & 3.24 \end{bmatrix} + \begin{bmatrix} 1.96 & 0 & -0.28 \\ 0 & 0 & 0 \\ -0.28 & 0 & 0.04 \end{bmatrix} + \begin{bmatrix} 5.76 & -9.6 & -10.1 \\ -9.6 & 16 & 16.8 \\ -10.1 & 16.8 & 17.64 \end{bmatrix}$$

$$+ \begin{bmatrix} 19.36 & 0 & -0.88 \\ 0 & 0 & 0 \\ -0.88 & 0 & 0.04 \end{bmatrix} + \begin{bmatrix} 0.16 & -2.4 & 1.12 \\ -2.4 & 36 & -16.8 \\ 1.12 & -16.8 & 7.84 \end{bmatrix} + \begin{bmatrix} 2.56 & -6.4 & -6.08 \\ -6.4 & 16 & 15.2 \\ -6.08 & 15.2 & 14.44 \end{bmatrix}$$

$$+ \begin{bmatrix} 12.96 & 0 & 15.12 \\ 0 & 0 & 0 \\ 15.12 & 0 & 17.64 \end{bmatrix} + \begin{bmatrix} 5.76 & 0 & 9.12 \\ 0 & 0 & 0 \\ 9.12 & 0 & 14.44 \end{bmatrix} + \begin{bmatrix} 31.36 & -16.8 & 6.72 \\ -16.8 & 9 & -3.6 \\ 6.72 & -3.6 & 1.44 \end{bmatrix}$$

$$+ \begin{bmatrix} 21.16 & 13.8 & 10.12 \\ 13.8 & 9 & 6.6 \\ 10.12 & 6.6 & 4.84 \end{bmatrix} \Rightarrow \Sigma = \begin{bmatrix} 12.04 & 0.5 & 3.28 \\ 0.5 & 12 & 2.9 \\ 3.28 & 2.9 & 8.16 \end{bmatrix} = U\Lambda U^T$$

$$\Rightarrow \Lambda \approx \text{diag}(15.23, 11.52, 5.45), \quad U \approx \begin{bmatrix} -0.63 & 0.67 & -0.40 \\ -0.57 & -0.75 & -0.35 \\ -0.53 & 0.008 & 0.85 \end{bmatrix}$$

$\Rightarrow$ 3 principle axes are $u_1 = \begin{bmatrix} -0.63 \\ -0.57 \\ -0.53 \end{bmatrix}$, $u_2 = \begin{bmatrix} 0.67 \\ -0.75 \\ 0.008 \end{bmatrix}$, $u_3 = \begin{bmatrix} -0.40 \\ 0.35 \\ 0.85 \end{bmatrix}$

b.

x_i	(1,2,3)	(4,8,5)	(3,12,9)	(1,8,5)	(5,14,2)	(7,4,1)	(9,8,9)	(3,8,1)	(11,5,6)	(10,11,7)
$u_1^T x$	-3.36	-9.73	-13.5	-7.84	-12.19	-7.22	-15	-6.98	-12.96	-16.28
$u_2^T x$	-0.81	-3.28	-6.92	-5.29	-7.13	1.7	0.1	-3.98	3.67	-1.49
$u_3^T x$	1.45	-0.15	2.25	1.05	-5.2	-3.35	1.25	-3.15	-1.05	-1.9

5.

C $\hat{x}_i^{(PCA)}$ on $span(u_1, u_2) = (u_1 u_1^T + u_2 u_2^T) x_i$

$$\Rightarrow \hat{X}^{(PCA)} = \begin{bmatrix} 0.846 & -0.143 & 0.339 \\ -0.143 & 0.887 & 0.296 \\ 0.339 & 0.296 & 0.281 \end{bmatrix} X$$

$$\Rightarrow X - \hat{X}^{(PCA)} = \begin{bmatrix} 0.154 & 0.143 & -0.339 \\ 0.143 & 0.113 & -0.296 \\ -0.339 & -0.296 & 0.719 \end{bmatrix} X$$

$$x_i - \hat{x}_i^{(PCA)} = (u_3 u_3^T) x_i$$

$$\Rightarrow X - \hat{X}^{(PCA)} = \begin{bmatrix} 0.157 & 0.137 & -0.337 \\ 0.137 & 0.12 & -0.294 \\ -0.337 & -0.294 & 0.723 \end{bmatrix} X$$

$$= \begin{bmatrix} -0.58 & 0.042 & -0.916 & -0.431 & 2.035 & 1.314 & -0.521 & 1.234 & 0.395 & 0.724 \\ -0.506 & 0.036 & -0.799 & -0.376 & 1.775 & 1.146 & -0.454 & 1.076 & 0.344 & 0.631 \\ 1.243 & -0.089 & 1.963 & 0.923 & -4.36 & -2.816 & 1.115 & -2.643 & -0.845 & -1.55 \end{bmatrix}$$

$$\Rightarrow \sum_{i=1}^{10} \| x_i - \hat{x}_i^{(PCA)} \|_2 = 20.639$$