Data Science

Prof M.-S. Chen
Prof M.-L Lo

TA Office ours

- Every Friday AM11:00-12:00 博理603
- Contact
  - email: datascience2020fall@gmail.com
  - FB QA **discuss community**

DS HW6 - NN

Outline

- Problems
- Training
- Report
- Submission

Outline

- **Problems**
- Training
- Report
- Submission

Problem 1 (40%) Basic Neural Networks

- Please train and test neural networks for establishing the relationship between input (x1, x2, x3, x4) and output (y1, y2) by using the program. You may try to investigate the effects of changing parameters such as learning rates, # of hidden units, # of hidden layers, moment, etc. You may show the performance (eg. MSE and running time) of the NN model you built with different parameters. Please predict the output (y1, y2) in following table based on your best established NN model. The dataset is shown in DS_hw6_p1.csv.

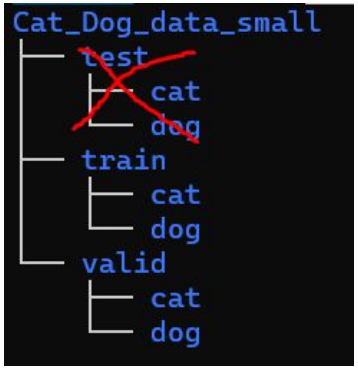
  * First 280 samples are training data, the remaining are testing data

  * Do not use testing data for training

Problem 2 (60%) Convolutional Neural Networks

- In problem 2 we need to train a NN to recognize cat&dog. You need to use the CNN layer module in your model now.
- Please use the dataset we provide.

* Dataset tree:



* Do not use valid data for training

Dataset

- ➔ Please use the dataset we provided.
- ➔ Dataset Link [link].

Outline

- Problems
- **Training**
- Report
- Submission

Learning Resources 😉

- ➔ Resources for problem1 : blog [link] 👍
- ➔ Example code for problem2 : github link [link] 👍
  - ◆ python3.6
  - ◆ avaliable packages: see requirements.txt
- ➔ What you need to do
  - ◆ tune parameters
  - ◆ modify your models

Outline

- Problems
- Training
- **Report**
- Submission

Problem 1 (40%) Basic Neural Networks

Please follow problems below and write down the answers in your report

1. Any prepocessing? (5%)
2. How you get the best model ? (5%)
3. State all the parameters you need for training (learning rates, epochs, weight decay, moment, etc.) (5%)
4. Show the structure of your best model (hint : print(model)) (5%)
5. Write down the prediction of testing data (y1, y2) (list in a table such as Table 1) (5%)
6. Plot the learning curve (MSE) (5%)

* hw6_p1.py (10%)

* if you do not apply a NN model, you won't get any point

Problem 1 (40%) Basic Neural Networks (Cont'd)

Table 1

No.	x_1	x_2	x_3	x_4	y_1	y_2
281	10.8	0.7	0.7	4.8		
282	5.4	12.0	-1.6	12.6		
283	-5.0	-12.1	21.4	-9.1		
284	-10.2	11.5	-4.3	0.8		
285	-0.9	2.1	-4.8	-2.7		
286	-1.0	-14.1	-19.2	0.5		
287	-10.3	-1.1	-5.2	6.8		
288	2.4	5.7	8.1	-6.0		
289	-25.9	-10.5	-1.5	-6.4		
290	-2.9	12.2	11.1	9.6		
291	11.7	13.3	-8.1	9.6		
292	17.2	-1.1	8.6	-0.8		
293	-8.4	-15.3	4.6	-4.2		
294	-12.1	2.0	11.7	-0.3		
295	1.3	-14.1	-4.3	7.8		
296	-13.3	-1.7	11.7	-5.7		
297	1.2	-7.0	-16.6	-1.2		
298	-8.1	4.6	7.9	12.9		
299	-7.3	-1.0	0.8	11.8		
300	10.3	-16.0	3.2	2.6		

Problem 2 (60%) Convolution Neural Networks

Please follow problems below and write down the answers in your report

1. State all the parameters you need for training (learning rates, epochs, weight decay, moment, etc.) (5%)
2. Show the structure of your best model (hint : print(model)) (5%)
3. Train/Valid accuacy of your model (5%)
4. Plot the learning curve (CrossEntropyLoss) (10%)

ex:

Problem 2 (60%) Convolution Neural Networks (Cont'd)

Reproduce part

1. **Simple baseline (25%)**

   valid set accuacy : 70%

   test set accuacy : 70% (privacy)

2. **Strong baseline (10%)**

   valid set accuacy : 75%

   test set accuacy : 75% (privacy)

*** hw6_p2.py (10%)**

*** if you do not apply a NN model, you wont get any point**

problem2 reproduce

- Environment :
  - python3.6
  - pytorch==1.4.0
  - avaliable packages: see requirements.txt
- TA will run :

```
> python3 hw6_p2.py --path <test_imgs_dir> --out_csv <csv_path>
```

Dir of images

use glob to read all image in the dir.

Output CSV format

should be same as p2/example_pred.csv

Label:

Cat:0
Dog:1

Outline

- Problems
- Training
- Report
- **Submission**

formate rules

- ⭐formate & submission ⭐
  - ➔ files/dir tree formate
    - hw6_<student_id>/
      - report.pdf
      - hw6_p1.py
      - hw6_p2.py
      - other_files (models pth)

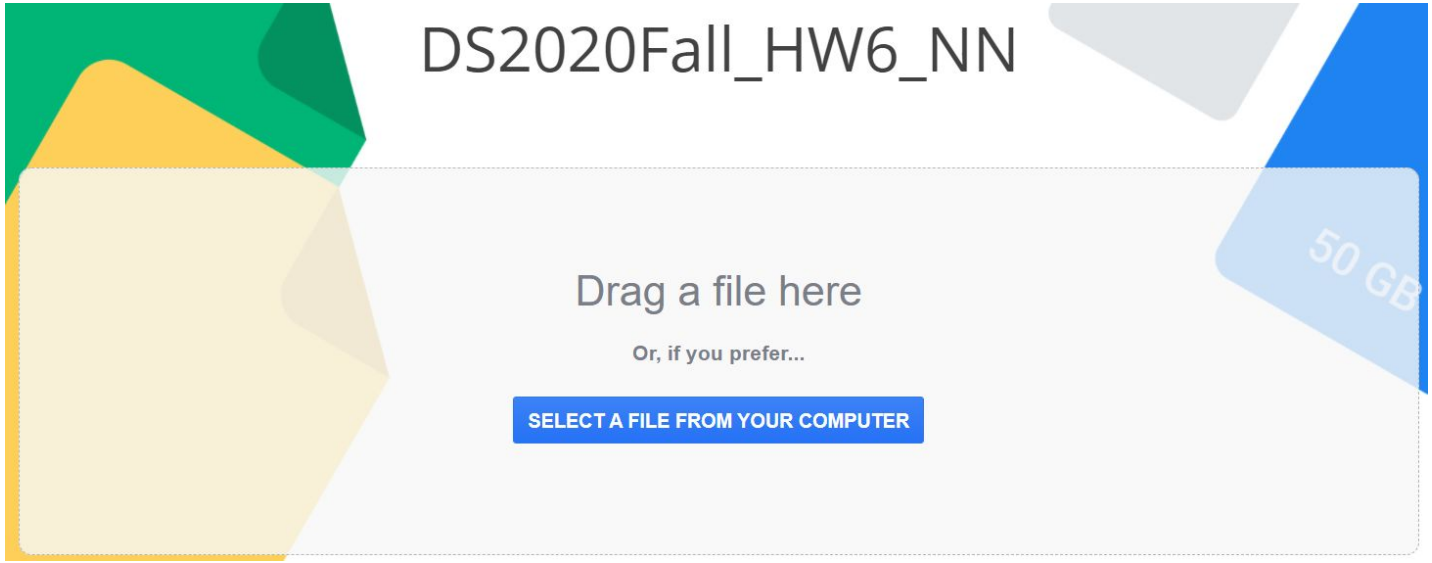
      (your trained model weight .pth files should include)

  - ➔ compress dir to hw6_<student_id>.zip

submission

- Deadline: **2020/12/31 Thurs. 23:59**
  - ➔ submission place: [link]

(no limited file size) (same filename will be replaced by newest one)



punishments rules

- late submission : you will get **0** points :<
- file formate error : **-10**
- missing files : **-10**
- execute error : **-5**

程式規定

- 程式語言: **Python3.6**
- 不熟悉程式語言的同學, 可在 TA hour 來詢問或寄信
  - TA hours: Every Friday AM11:00-12:00 博理603