AI 應用於醫療分析

107302002 金融四 張致中

資料內容:

- 資料來源: Single Proton Emission Computed Tomography heart diagnosis dataset
- 共有267筆資料，其中有55筆是正常，212筆是不正常，約1:4
- Data有22筆binary 的attributes, Label也是 binary資料
- 其中0為正常，1為不正常
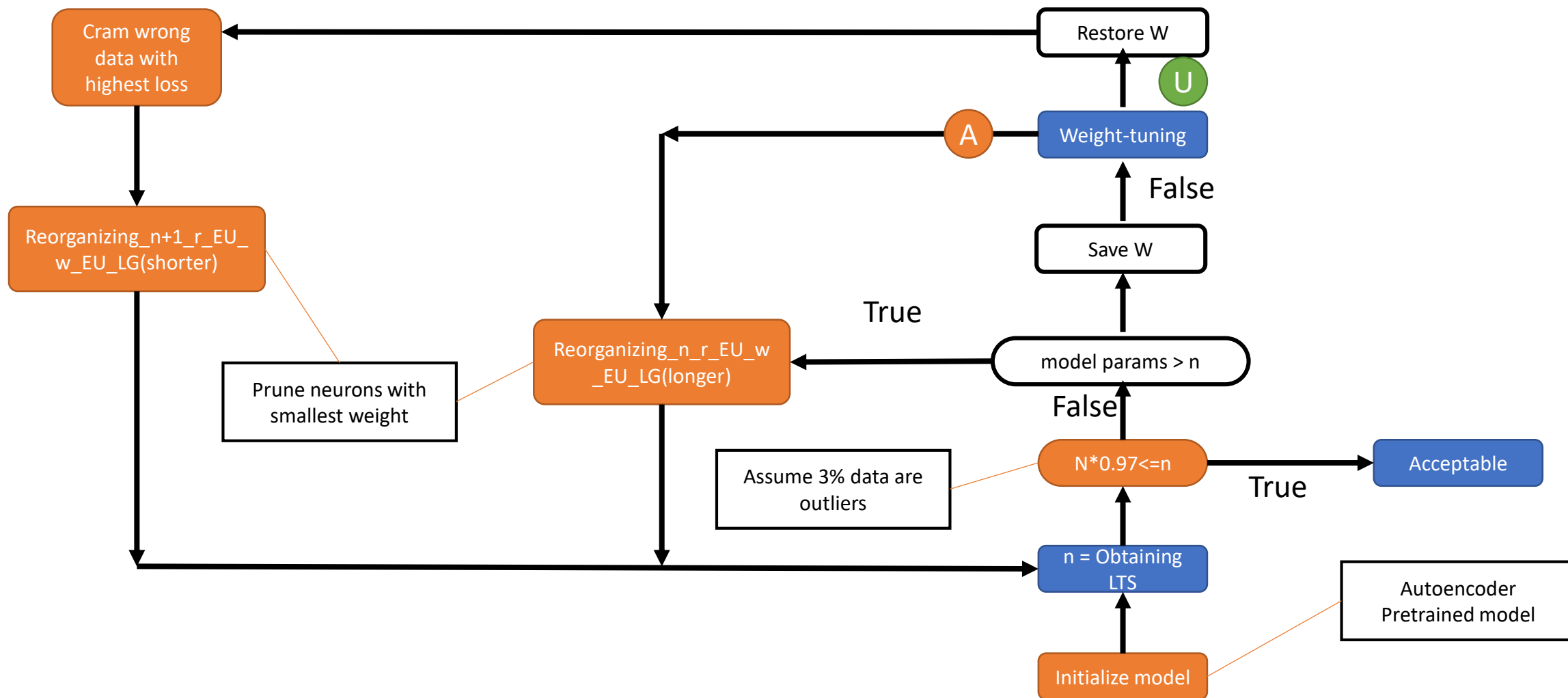- 本次實驗按8:2分成training data和validation data，隨機產生20組資料。
- 不平均+離群子

環境配置

硬體與**python**版本:
- CPU: R5 2400g
- Ram: 32G
- GPU: RTX 3060
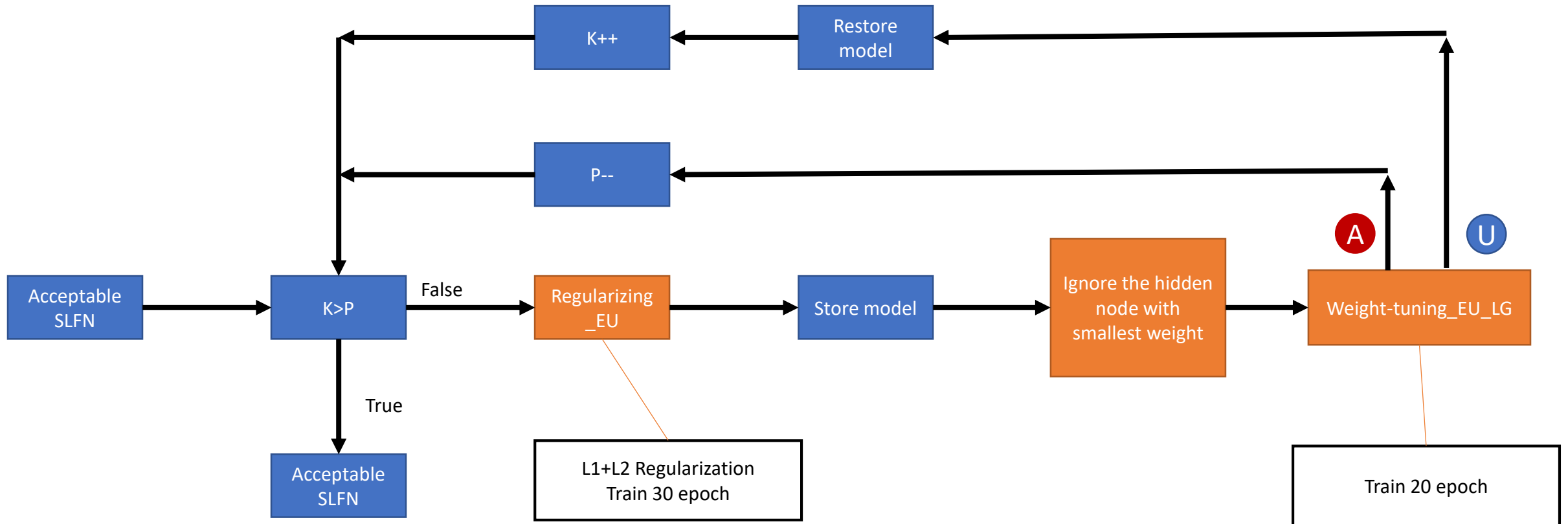- IDE: Vscode
- Python: 3.7.6

套件:
- Pytorch
- Sklearn
- Pandas
- Itertools
- Random
- Copy
- Logging
- OS
- Time

新型學習演算法:

Reorganizing_ALL_r_EU_w_EU_LG

模型參數

Input num

Output: as close as input

bottleneck

encoder

decoder
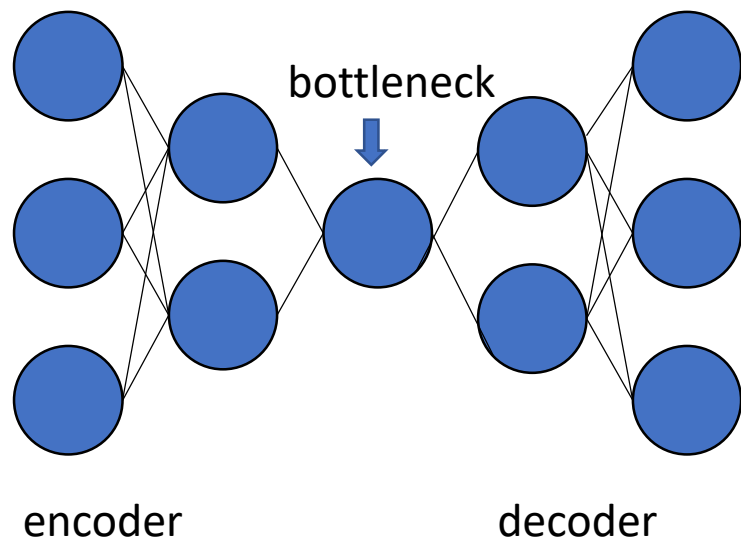
使用這部分的模型

Optimizer:
AdamW

Activation function:
ReLU

Loss function:
Focal loss

處理樣本不均

$$FL = \begin{cases} -\alpha(1-p)^\gamma log(p), & if \quad y=1 \\ -(1-\alpha)p^\gamma log(1-p), & if \quad y=0 \end{cases}$$

Focal loss 論文

Validate the proposed learning mechanism
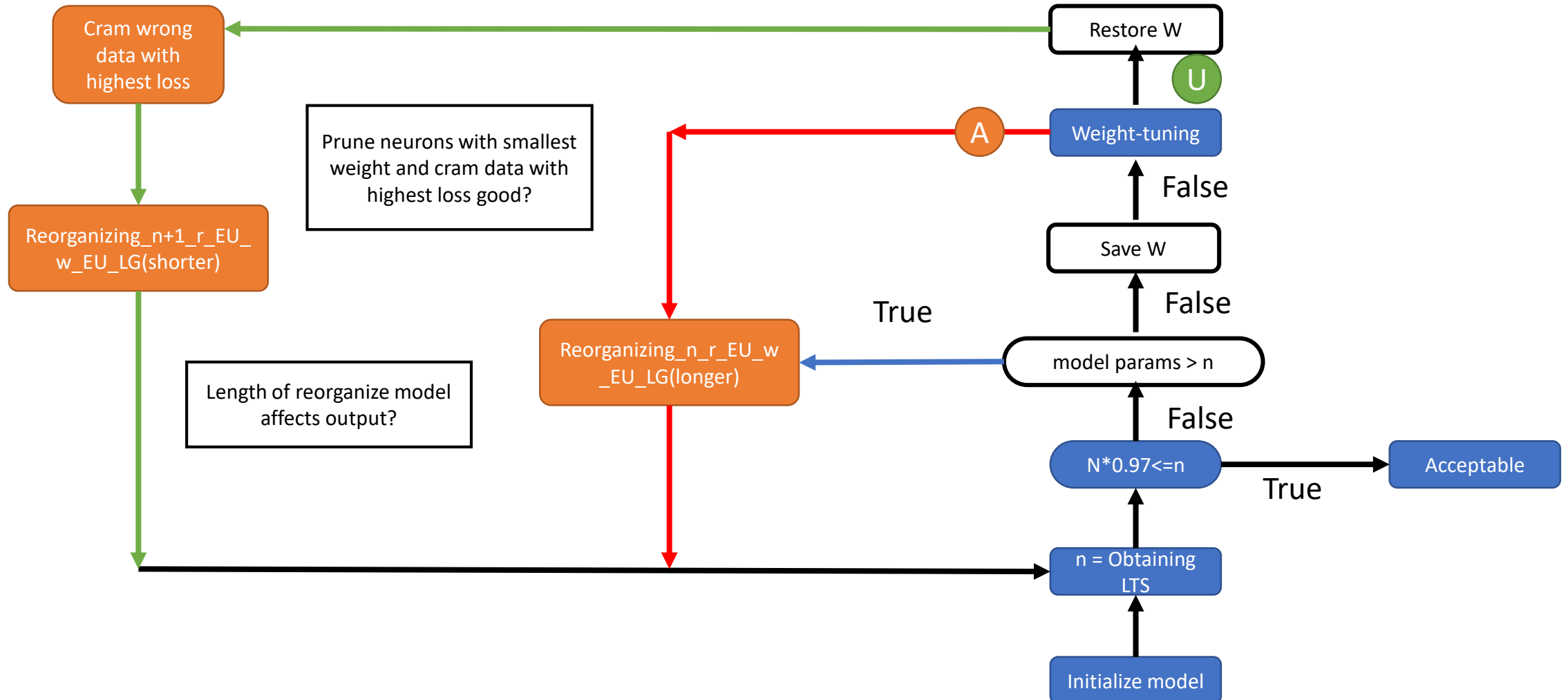
Four versions

Version	Reorganizing Module	Cramming and Pruning
Reorganize – 10	Reorganizing(10, 30)	Cram highest loss, prune smallest weight
Reorganize – 50	Reorganizing(50, 150)	Cram highest loss, prune smallest weight
Reorganize – 100	Reorganizing(100, 300)	Cram highest loss, prune smallest weight
Cram min Prune max	Reorganizing(10, 30)	Cram lowest loss, prune biggest weight

Reorganize 10 (baseline)

Set	Green	Blue	Red
1	6	5	1
2	0	0	1
3	8	6	1
4	3	0	1
5	5	12	3
6	0	1	1
7	0	0	1
8	13	0	9
9	8	6	8
10	4	1	1
11	0	2	1
12	3	18	7
13	3	0	1
14	8	6	8
15	4	4	1
16	13	0	1
17	18	0	9
18	0	1	1
19	13	0	9
20	6	1	1
AVG	5.75	3.15	3.3
STD	5.087976	4.618171	3.348134

Reorganize 50

- Green Average is the highest -> cram more

Set	Green	Blue	Red
1	1	1	1
2	3	2	1
3	0	2	1
4	2	4	1
5	1	0	1
6	19	0	9
7	15	0	9
8	0	1	1
9	13	14	1
10	1	0	1
11	35	0	9
12	4	8	8
13	12	1	1
14	7	7	8
15	6	7	1
16	6	0	1
17	11	6	8
18	5	0	1
19	4	7	7
20	9	13	1
AVG	7.7	3.65	3.55
STD	8.167619	4.315959	3.499643

Reorganize 100

- Blue Average is the lowest -> no overfitting

- Maybe longer reorganize helps pruning

Set	Green	Blue	Red
1	2	2	1
2	25	0	9
3	0	2	1
4	0	1	1
5	0	1	1
6	8	2	1
7	1	0	1
8	11	7	1
9	8	0	1
10	7	0	1
11	0	1	1
12	14	0	9
13	4	0	1
14	0	1	1
15	7	1	1
16	14	4	1
17	3	6	1
18	18	0	9
19	14	0	9
20	0	1	1
AVG	6.8	1.45	2.6
STD	7.03278	1.961505	3.2

Cram min Prune max

- Cramming has less effect on model loss, so the number of n come from LTS module increase very slow.

Set	Green	Blue	Red
1	12	10	1
2	3	4	1
3	1	10	7
4	6	9	1
5	5	10	7
6	10	2	1
7	3	12	8
8	11	8	8
9	4	3	1
10	8	9	6
11	6	8	8
12	3	8	7
13	3	5	1
14	12	0	9
15	1	8	1
16	4	8	1
17	2	8	1
18	6	8	7
19	6	9	8
20	0	2	1
AVG	5.3	7.05	4.25
STD	3.565109	3.153966	3.299621

Green Path: Cramming

- Prune more, cram more

Set	Reorganize 10	Reorganize 50	Reorganize 100	Cram min Prune max
1	6	1	2	12
2	0	3	25	3
3	8	0	0	1
4	3	2	0	6
5	5	1	0	5
6	0	19	8	10
7	0	15	1	3
8	13	0	11	11
9	8	13	8	4
10	4	1	7	8
11	0	35	0	6
12	3	4	14	3
13	3	12	4	3
14	8	7	0	12
15	4	6	7	1
16	13	6	14	4
17	18	11	3	2
18	0	5	18	6
19	13	4	14	6
20	6	9	0	0
AVG	5.75	7.7	6.8	5.3
STD	5.087976	8.167619	7.03278	3.565109

Blue Path: Big param

Cram min Prune max:
需要花較多時間才能夠達到Pruning門檻->
沒有頻繁的去Pruning

Set	Reorganize 10	Reorganize 50	Reorganize 100	Cram min Prune max
1	5	1	2	10
2	0	2	0	4
3	6	2	2	10
4	0	4	1	9
5	12	0	1	10
6	1	0	2	2
7	0	0	0	12
8	0	1	7	8
9	6	14	0	3
10	1	0	0	9
11	2	0	1	8
12	18	8	0	8
13	0	1	0	5
14	6	7	1	0
15	4	7	1	8
16	0	0	4	8
17	0	6	6	8
18	1	0	0	8
19	0	7	0	9
20	1	13	1	2
AVG	3.15	3.65	1.45	7.05
STD	4.618171	4.315959	1.961505	3.153966

Red Path: Weight-tune

Cram min Prune max:
需要花較多時間才能夠達到Pruning門檻->
沒有頻繁的去Pruning

Set	Reorganize 10	Reorganize 50	Reorganize 100	Cram min Prune max
1	1	1	1	1
2	1	1	9	1
3	1	1	1	7
4	1	1	1	1
5	3	1	1	7
6	1	9	1	1
7	1	9	1	8
8	9	1	1	8
9	8	1	1	1
10	1	1	1	6
11	1	9	1	8
12	7	8	9	7
13	1	1	1	1
14	8	8	1	9
15	1	1	1	1
16	1	1	1	1
17	9	8	1	1
18	1	1	9	7
19	9	7	9	8
20	1	1	1	1
AVG	3.3	3.55	2.6	4.25
STD	3.348134	3.499643	3.2	3.299621

Accuracy

Average:
[55+(267-55)*0.784]/267 = 82.8%

Best:
[55+(267-55)*0.871]/267 = 89.7%

Original paper:
Rule based model: 84%

Maybe…

Cram highest loss ->
Reach the learning goal too fast?

Prune smallest weight ->
Prune the well learned node?

Set	Reorganize 10		reorganize 50		reorganize 100		Cram min prune max	
1	0.5548	0.5455	0.8051	0.7159	0.5236	0.5085	0.8029	0.7528
2	0.7963	0.7912	0.7898	0.8153	0.8501	0.8224	0.8656	0.8068
3	0.8031	0.7614	0.7827	0.8011	0.6023	0.6378	0.7874	0.7443
4	0.7876	0.8452	0.5236	0.4929	0.5032	0.5369	0.8727	0.7528
5	0.642	0.5625	0.5412	0.5767	0.7944	0.8224	0.875	0.8707
6	0.8031	0.7543	0.8478	0.8608	0.8142	0.7216	0.7785	0.7628
7	0.5884	0.6151	0.8478	0.8452	0.8006	0.7457	0.8412	0.8224
8	0.8299	0.8011	0.7989	0.7997	0.5548	0.5284	0.8548	0.8537
9	0.8255	0.669	0.8168	0.7301	0.8051	0.7301	0.8433	0.8395
10	0.5236	0.608	0.5278	0.4858	0.6025	0.5085	0.8163	0.6392
11	0.8121	0.7088	0.8703	0.8295	0.8031	0.7614	0.8212	0.8466
12	0.8703	0.777	0.8389	0.7926	0.8189	0.7756	0.801	0.8551
13	0.7987	0.777	0.5593	0.554	0.6241	0.5625	0.8567	0.7685
14	0.8433	0.8082	0.8793	0.7855	0.5123	0.6605	0.7834	0.7543
15	0.821	0.6918	0.5323	0.554	0.8053	0.7173	0.8163	0.6776
16	0.5548	0.4162	0.7915	0.7472	0.8123	0.7741	0.7942	0.7784
17	0.8525	0.8537	0.801	0.7074	0.51	0.5312	0.8793	0.8011
18	0.5614	0.4915	0.1924	0.2543	0.8391	0.8849	0.8367	0.8466
19	0.8278	0.7926	0.8299	0.8693	0.8278	0.8139	0.8612	0.6932
20	0.6108	0.6236	0.8858	0.794	0.8102	0.7614	0.841	0.8153
AVG	0.73535	0.694685	0.72311	0.700565	0.710695	0.690255	0.831435	0.784085
STD	0.120038	0.120612	0.176686	0.157994	0.131668	0.118312	0.031799	0.062158

Training Time

- Model initialize後loss值低但accuracy不高

- 也許可以加入正確率做為 learning goal

Set	reorganize 10	reorganize 50	reorganize 100	cram min prune max
1	438.883	406.602	1615.752	768.053
2	9.991	959.892	10145.78	237.832
3	550.653	248.09	521.86	401.002
4	116.621	1477.942	178.08	505.522
5	648.772	101.18	204.48	653.852
6	18	3526.704	3413.732	508.772
7	9.191	2842.004	176.27	651.063
8	543.444	88.22	6514.463	845.483
9	524.163	4813.063	2948.033	253.102
10	167.621	87.1	2741.074	619.523
11	35.63	6643.984	177.11	525.813
12	748.292	2192.693	5299.884	412.652
13	113.301	2337.464	1339.641	263.71
14	517.583	2528.103	177.32	472.444
15	281.572	2416.093	2981.403	324.772
16	475.344	1188.903	6388.833	426.372
17	660.294	3269.733	3166.002	350.052
18	17.97	901.061	7173.544	521.982
19	492.034	2260.302	5441.534	553.783
20	265.093	3789.183	177.5	36.7
AVG	331.7226	2103.916	3039.115	466.6242
STD	247.395	1696.107	2848.452	188.5243

Thanks for Listening!