Contents

Summary

Remove outliers

Class balance: same size for each class

KFold = 3

Use weight decay (http://stats.stackexchange.com/questions/70101/neural-networks-weight-change-momentum-and-weight-decay)

Raw data

Training samples x features = 61878 x 93

Number of test samples = 144368

Class 1:	1929
Class 4:	2691
Class 5:	2739
Class 7:	2839
Class 9:	4955
Class 3:	8004
Class 8:	8464
Class 6:	14135
Class 2:	16122

Remove outliers

Using PCA and Mahal distance

Threshold	Min size
150	1767
141	1749
170	1803
160	1776
155	1772

Threshold is based on this link http://sites.stat.psu.edu/~mga/401/tables/Chi-square-table.pdf

Class name	Number raw samples	Number *normal* samples	Number outlier samples
Class 1:	1929	1772	157
Class 4:	2691	2404	287
Class 5:	2739	2470	269
Class 7:	2839	2574	265
Class 9:	4955	4520	435
Class 3:	8004	7246	758
Class 8:	8464	7730	734
Class 6:	14135	12920	1215
Class 2:	16122	14634	1488

Using Mahal only

Not yet

Using ZScore or others

Not yet

Tackle class unbalance

Training

No class balance

Iteration number: 1000, Learning rate:0.001000, Hidden unit number: 25

Accuracy	LogLoss
0.4792	9.7814
0.0450	14.2861
0.2103	16.8889
---	---
0.2448	13.6521

Use all data, simple weight, 3-fold

Class balance: incl. Outliers vs excl. Outliers

Data from ..\Data\Classes\ Num samples per class: 1929, min size	Data from ..\Data**RemoveOutliers**\ Num samples per class: 1772

Iteration number: 1000, Learning rate:0.001000, Hidden unit number: 25 Accuracy LogLoss 0.7216 0.5118 0.7330 0.4998 0.7470 0.4631 ---------- 0.7339 0.4916	Iteration number: 1000, Learning rate:0.001000, Hidden unit number: 25 Accuracy LogLoss 0.7492 0.4619 0.7578 0.4155 0.7601 0.4218 ---------- 0.7557 0.4331

Notes:

Excl. Outliers based on PCA and Mahal distance

Simple weight calculation, 3-fold, min size, Outliers threshold=155

Excl. Outliers: with Larger class size

Data from ..\Data\RemoveOutliers\ Num samples per class: 4000 Iteration number: 1000, Learning rate:0.001000, Hidden unit number: 25 Accuracy LogLoss 0.3403 5.3085 0.7606 0.4282 0.4138 4.6889 ---------- 0.5049 3.4752 numSamples from each class <= max samples allow different size of each class	Data from ..\Data\RemoveOutliers\ Num samples per class: 4000 Iteration number: 1000, Learning rate:0.001000, Hidden unit number: 25 Accuracy LogLoss 0.3152 4.5638 0.4658 4.3501 0.2462 6.2646 ---------- 0.3424 5.0595 Each class contributes the same number of samples. Class with less samples is duplicated

Notes: Simple weight calculation, 3-fold, Outliers threshold=155

KFold

Data from ..\Data\RemoveOutliers\ Num samples per class: 1772, K=3 Iteration number: 1000, Learning rate:0.001000, Hidden unit number: 25 Accuracy LogLoss 0.7533 0.4494 0.7522 0.4438 0.7561 0.4277 ---------- 0.7539 0.4403	Data from ..\Data\RemoveOutliers\ Num samples per class: 1772 Iteration number: 1000, Learning rate:0.001000, Hidden unit number: 25, 5-fold Accuracy LogLoss 0.7524 0.4352 0.7470 0.4501 0.7615 0.4609 0.7389 0.4560 0.6191 2.2513 ---------- 0.7238 0.8107

K=4 or 7: result is worser

Data from ..\Data\RemoveOutliers\
Num samples per class: 1772
Iteration number: 1000, Learning rate:0.001000, Hidden unit number: 25, 2-fold
Accuracy LogLoss

0.7448 0.4478
0.7489 0.4350

0.7469 0.4414

Weight decay

Data from ..\Data\RemoveOutliers\ Num samples per class: 1772 Iteration number: 1000, Learning rate:0.001000, Hidden unit number: 25, 3-fold Lamda = 1.9 Accuracy LogLoss 0.7589 0.3423 0.7409 0.3726 0.7419 0.3725 --- 0.7472 0.3625	Data from ..\Data\RemoveOutliers\ Num samples per class: 1772 Iteration number: 1000, Learning rate:0.001000, Hidden unit number: 25, 3-fold Lamda = <=1.5 Accuracy LogLoss 0.7511 0.3910 0.7565 0.3732 0.7482 0.3868 --- 0.7519 0.3837

Notes: low log loss but also lower accuracy. Should check this

2 hidden layers

Not yet

Reduce features

PCA without ZScore: poor result

Accuracy LogLoss
0.5991 0.6373
0.6279 0.6888
0.6146 0.7211

0.6139 0.6824

NN

Linear function at the output layer