deepGTA training log:

2017-03-07

Sigmoid activation function. X Not converged. Weight highly biased.

Change to relu function. √ Converged.

Data set: hsq = 0.6

Test1: 2017-03-07  
network setting: 1000x1000x200 relu

dataset : 4000 train 1000 valid 1000 test

result:

Step: 5000

Training Data Eval:

Num samples: 4000 R\_squared: 0.9394

Validation Data Eval:

Num samples: 1000 R\_squared: -1.9654

Test Data Eval:

Num samples: 1000 R\_squared: -2.1086

Conclusion: overfitting.

Test2: 2017-03-07

network setting: 1000x200x20 relu  
dataset : 8000 train 1000 valid 1000 test

Step:5000

Training Data Eval:

Num samples: 8000 R\_squared: 0.8709

Validation Data Eval:

Num samples: 1000 R\_squared: 0.3846

Test Data Eval:

Num samples: 1000 R\_squared: 0.4035

Conclusion: overfitting.

Test3: 2017-03-07

network setting: 1000x200x20 relu  
dataset : 20000 train 1000 valid 1000 test

Step:5000

Training Data Eval:

Num samples: 20000 R\_squared: 0.6872

Validation Data Eval:

Num samples: 1000 R\_squared: 0.5136

Test Data Eval:

Num samples: 1000 R\_squared: 0.5691

Conclusion: still overfitting, but better.

Test4: 2017-03-07

network setting: 1000x200x20 relu  
dataset : 30000 train 1000 valid 1000 test

Step:5000

Training Data Eval:

Num samples: 30000 R\_squared: 0.6510

Validation Data Eval:

Num samples: 1000 R\_squared: 0.5574

Test Data Eval:

Num samples: 1000 R\_squared: 0.5647

Conclusion: Nearly no overfitting

Test4: 2017-03-07

network setting: 1000x200x20 relu  
dataset : 100000 train 4000 valid 4000 test

Step:10000

Training Data Eval:

Num samples: 100000 R\_squared: 0.6118

Validation Data Eval:

Num samples: 4000 R\_squared: 0.6003

Test Data Eval:

Num samples: 4000 R\_squared: 0.5880

2017-03-08

Goal today: Make it converge in the small sample\_size as well as overcome the overfitting.

Add dropout.

Test1:

Dropout prob = [0.5,0.5]

Weight\_decay = 0,

Network: 1000x200x20

Dataset 4000 train, 1000 valid, 1000 test.

Training Data Eval:

Num samples: 5000 R\_squared: 0.7964

Validation Data Eval:

Num samples: 1000 R\_squared: 0.3706

Test Data Eval:

Num samples: 1000 R\_squared: 0.2839

Test2:

Dropoutporb = [0.8,0.8]

Weight\_decay = 0,

Network: 1000x200x20

Dataset 4000 train, 1000 valid, 1000 test

Training Data Eval:

Num samples: 5000 R\_squared: 0.9215

Validation Data Eval:

Num samples: 1000 R\_squared: 0.3755

Test Data Eval:

Num samples: 1000 R\_squared: 0.3796

Test3:

KEEP\_PROB = [0.9,0.9]

Weight\_decay = 0.0001,

Training Data Eval:

Num samples: 5000 R\_squared: 0.9571

Validation Data Eval:

Num samples: 1000 R\_squared: 0.4577

Test Data Eval:

Num samples: 1000 R\_squared: 0.4227