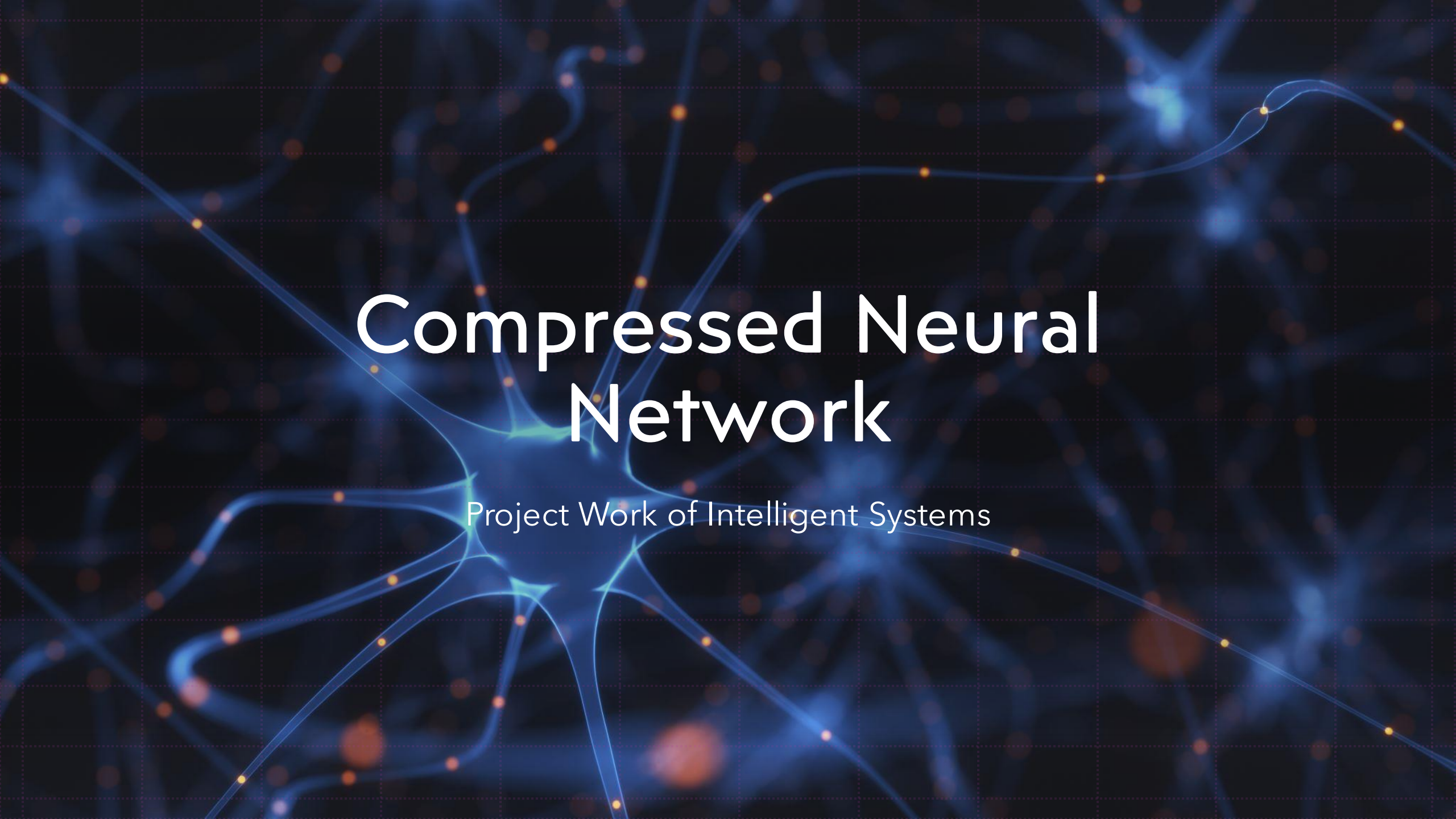


Compressed Neural Network

Project Work of Intelligent Systems

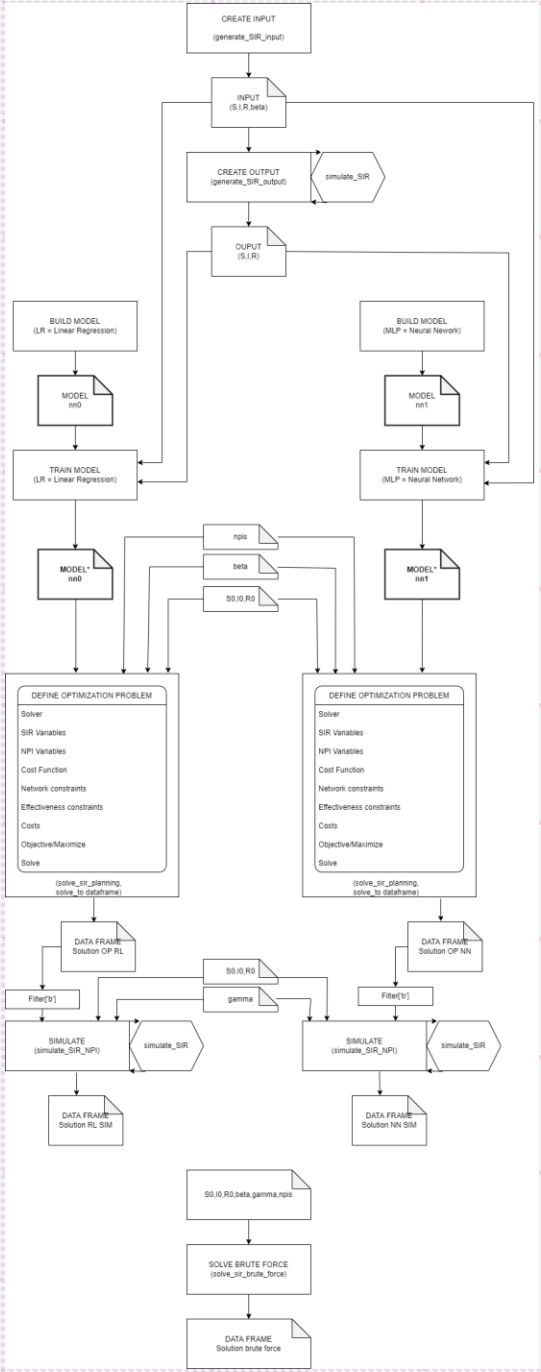
Epidemics

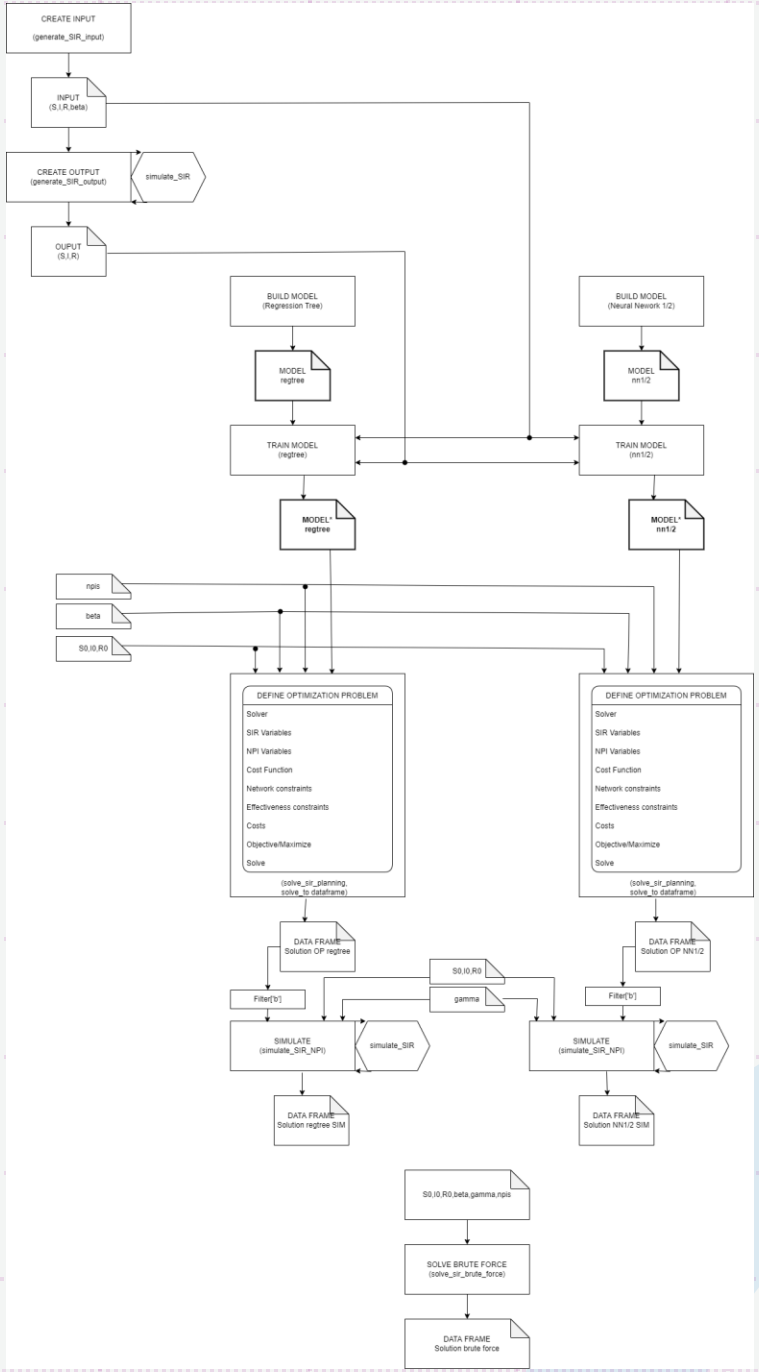
Rebuild pipelines

From
DecisioneTree and single NN
To
DecisionTree and multiple NN

Pipeline

Data
Train
Encode
Solve
Simulate
Brute Force





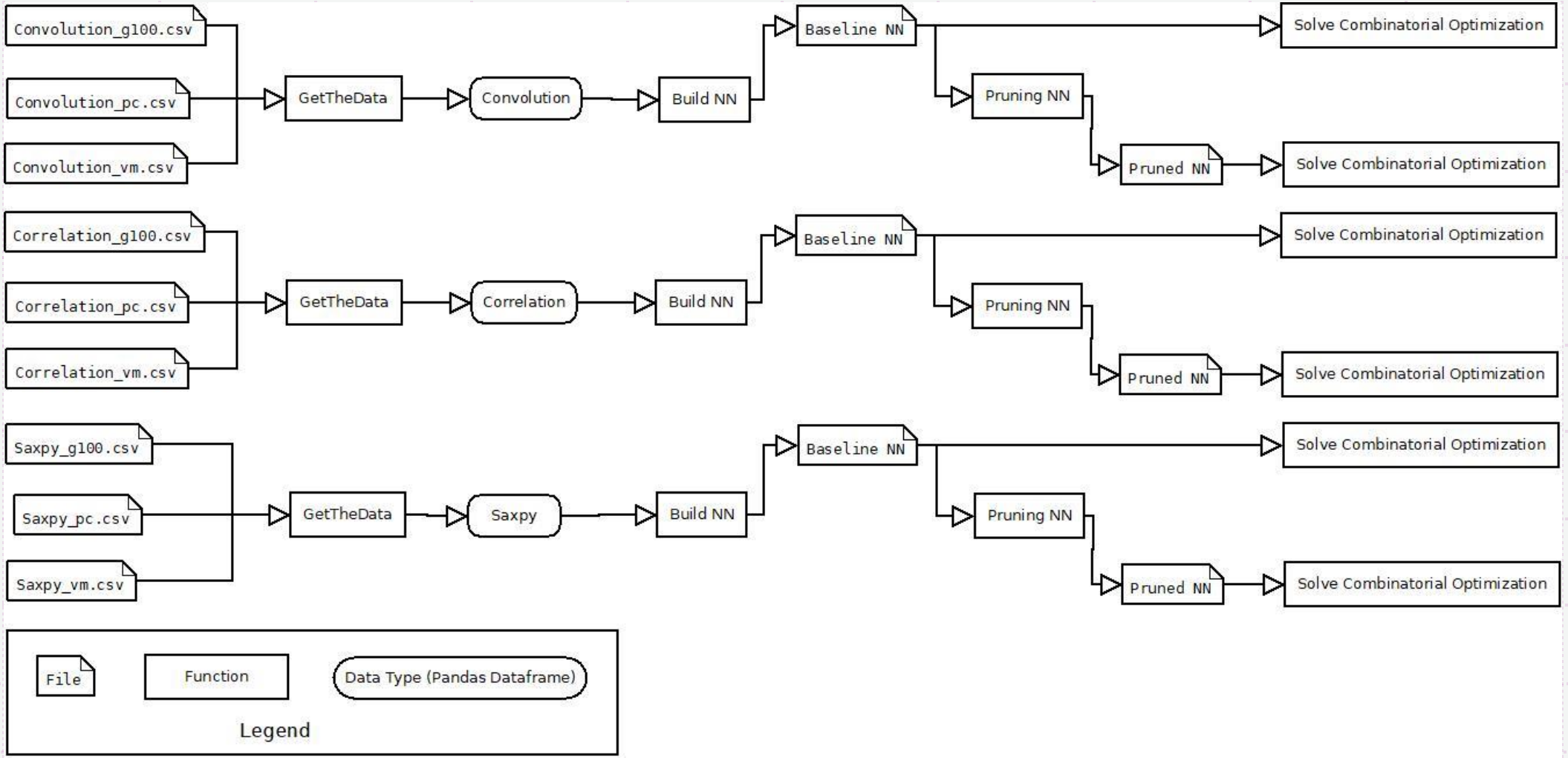
Epidemics

TensorFlow Model Optimization Toolkit

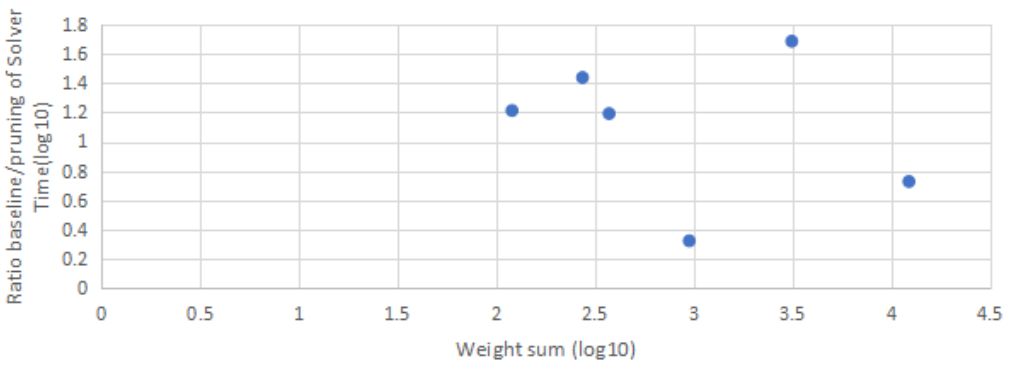
1. Post Training Quantization
    1. **Pro** : reduce size , most technique do not lose accuracy
    2. **Cons** : does not allow encoding in EML
2. Pruning
    1. **Pro** : does not lose accuracy , allows encoding in EML
    2. **Cons** : increase file size

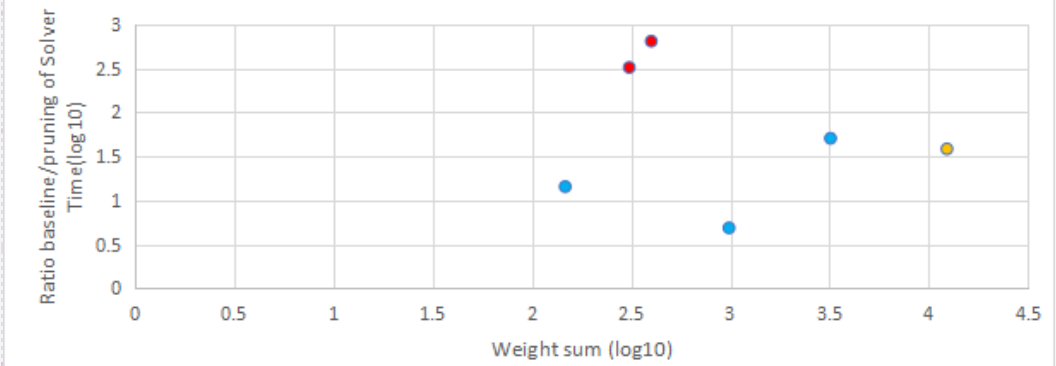
Transprecision Computing

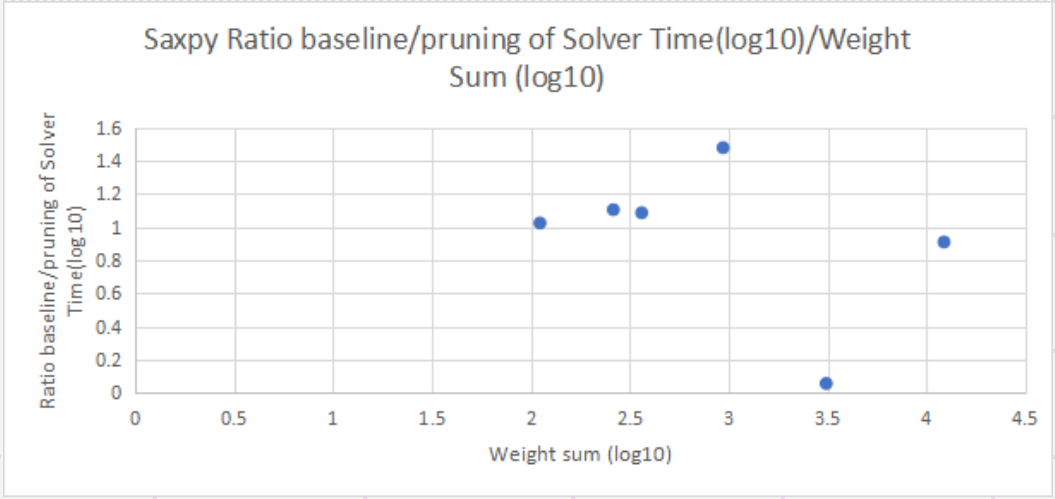
Pipeline



Result


Convolution Ratio baseline/pruning of Solver Time(log10) /Weight Sum (log10)
Ratio baseline/pruning of Solver Time(log10)
Weight sum (log10)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5


Correlation Ratio baseline/pruning of Solver Time(log10)/Weight Sum (log10)
Ratio baseline/pruning of Solver Time(log10)
Weight sum (log10)
0
0.5
1
1.5
2
2.5
3
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5


Saxpy Ratio baseline/pruning of Solver Time(log10)/Weight Sum (log10)
Ratio baseline/pruning of Solver Time(log10)
Weight sum (log10)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5