

Hyperparameter tuning

Network parameters:

- Number of layers
- Layer sizes
- Dropout: on/off
- Batch normalization:
on/off
- Nonlinearity

Training parameters:

- Learning rate
- Momentum

Usually finds better optima than when tuned by hand

Honest comparison with other methods in research

