Fast Training MLP Networks with Lo-Shu Data Sampling

Graphion LLC.

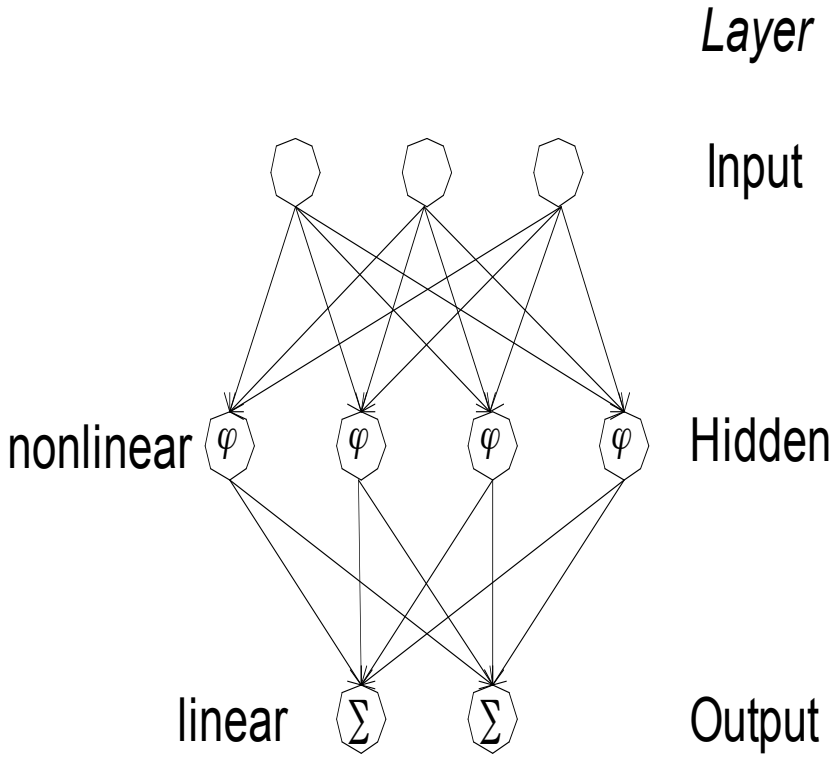
Hung-Han Chen, Ph. D.

Outline

- MLP & Backpropagation
- Training Curve
- Data Sampling
- Lo-Shu Square
- Simulation Results
- Conclusion
- Q & A

MLP

- Connectionism
- Feed Forward
- Nonlinear Activation
- Error Backpropagation
- Supervised Learning
- Least Square Algorithm
- Gradient Descents



MLP

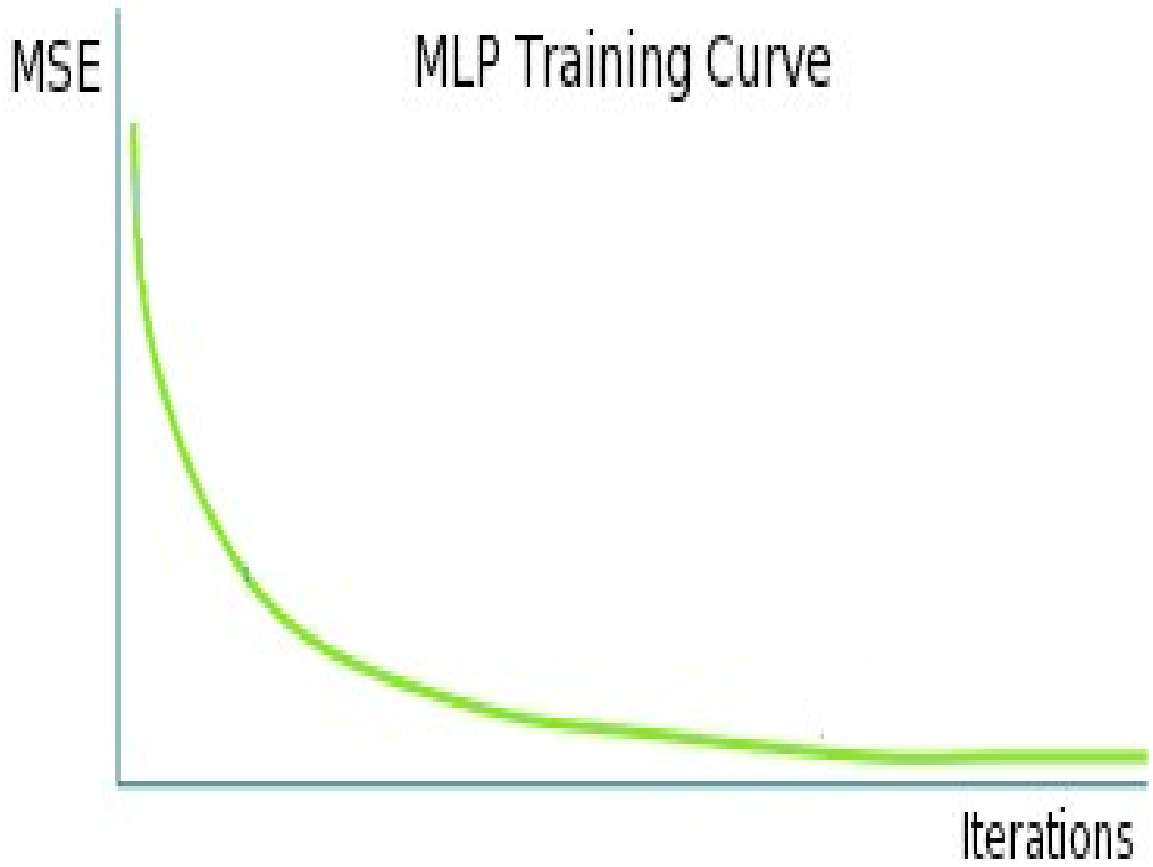
4 major concerns

- Big Scale/Rare Events
- Local Minima
- Generalization
  (Over Training)
- Slow Speed

Solutions

- NeuroSequences
  (FFC 2008, DC, USA)
- Retreat & Turn Search
  (ISNN 2008, Beijing, China)
- Monitoring Parameters
  (AIQED 2009, Cambridge, UK)
- Lo-Shu Data Sampling
  (AIQED 2009, Cambridge, UK)

Training Curve



Data Sampling

- A Process of selecting samples from the whole dataset.
- This can reduce the resource, i.e. time, used on model building through reducing the number of training samples.
- The training samples need to capture the real characteristics of whole dataset.

Lo-Shu Square

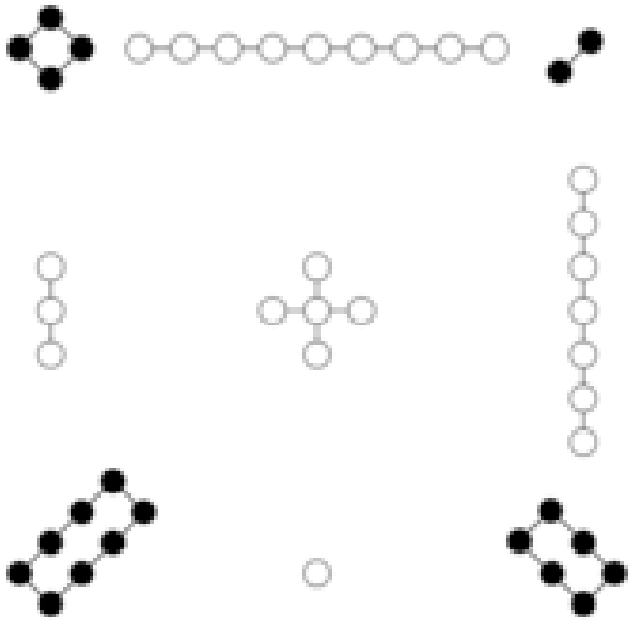


Fig. 2. Lo-Shu Square

{4, 9, 2} (戴九)
{8, 1, 6} (履一)
{4, 3, 8} (左三)
{2, 7, 6} (右七)
{9, 5, 1}
{3, 9, 7} (二四為肩)
{2, 5, 8}
{4, 5, 6} (六八為足)
{3, 1, 7} (五在中間)

Fig. 3. Nine groups from Lo-Shu Square

Simulation Results

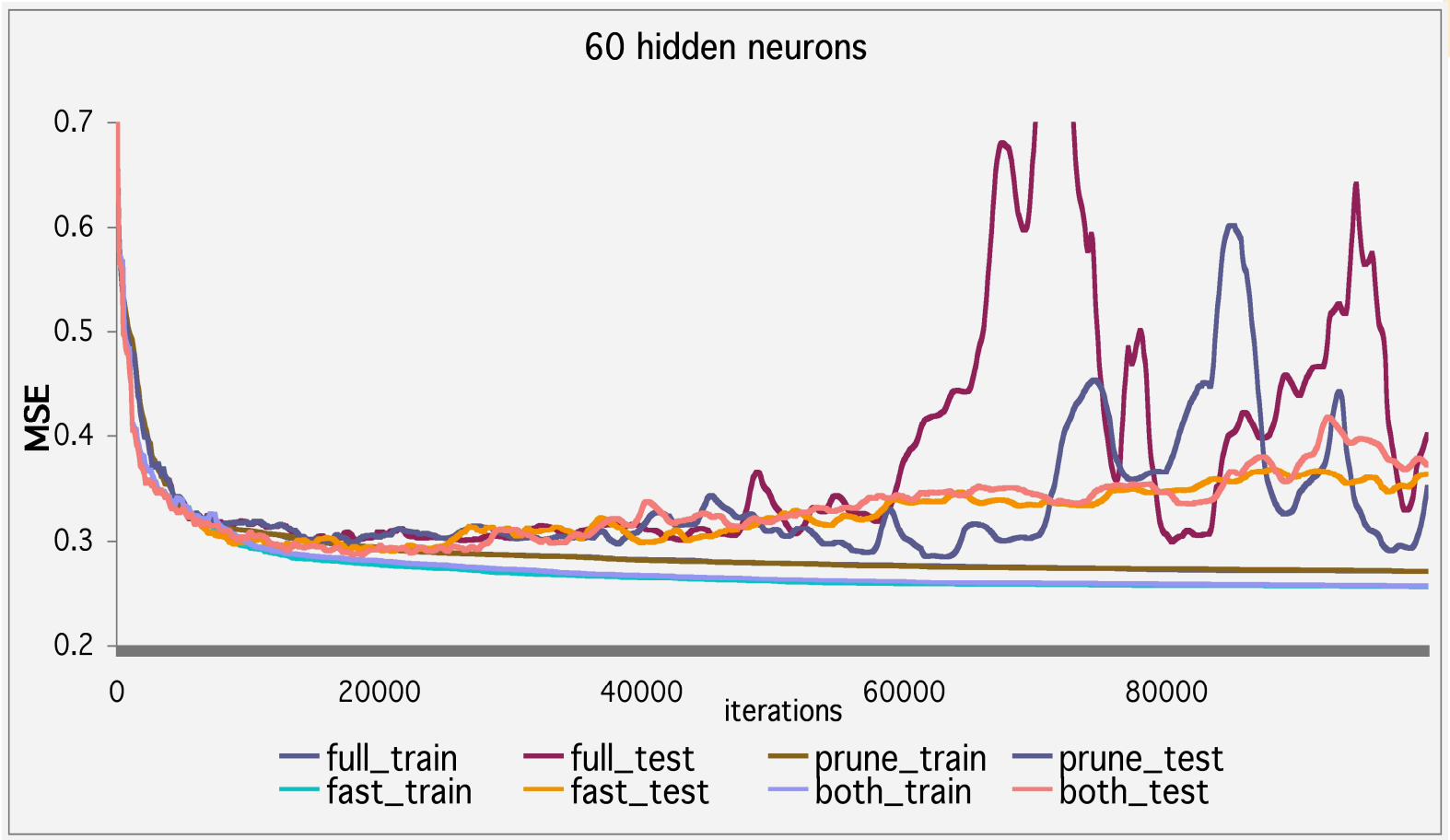


Fig. 4. 60 hidden neurons

Simulation Results

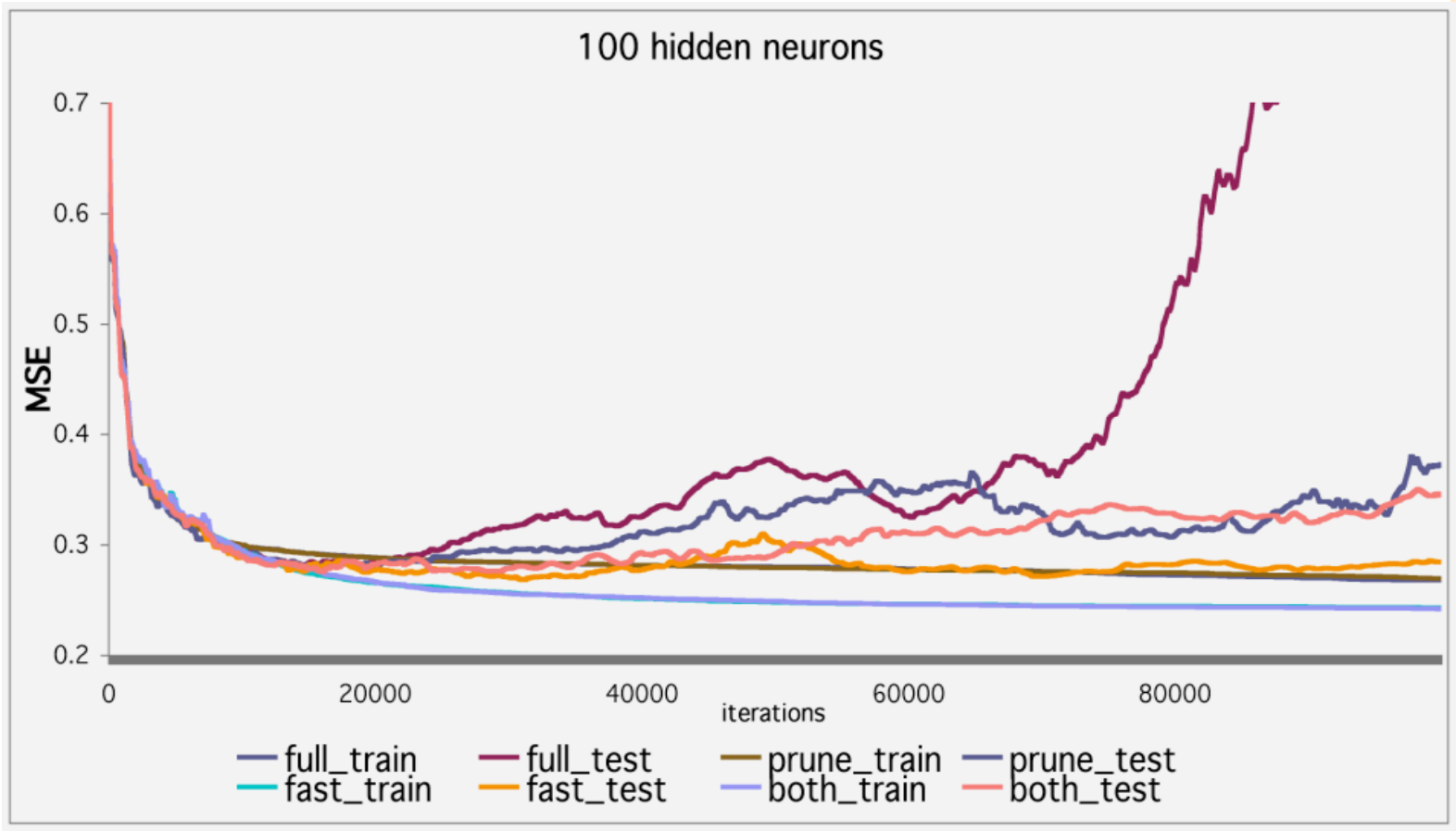


Fig. 5. 100 hidden neurons

Simulation Results

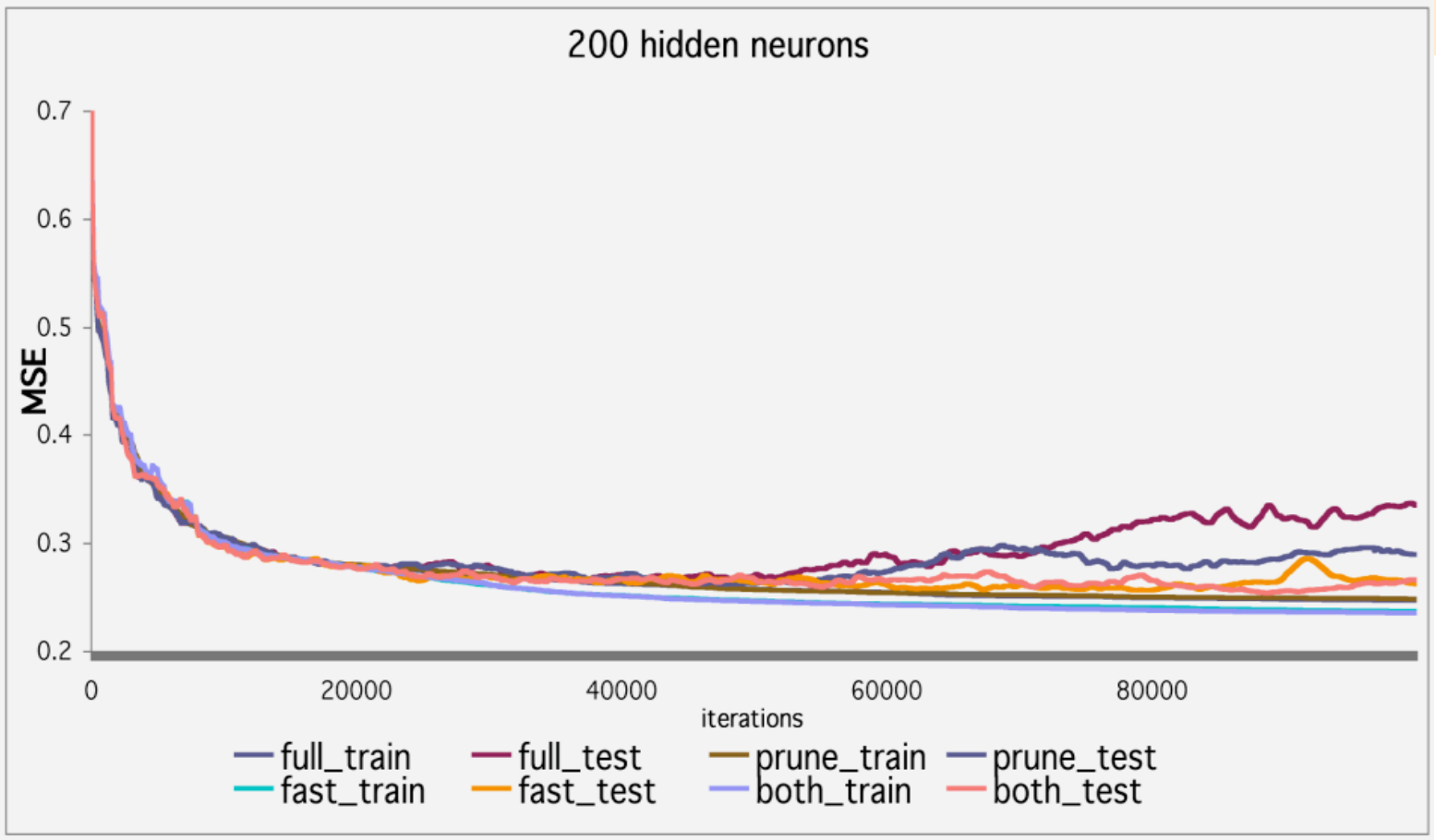


Fig. 6. 200 hidden neurons

Simulation Results

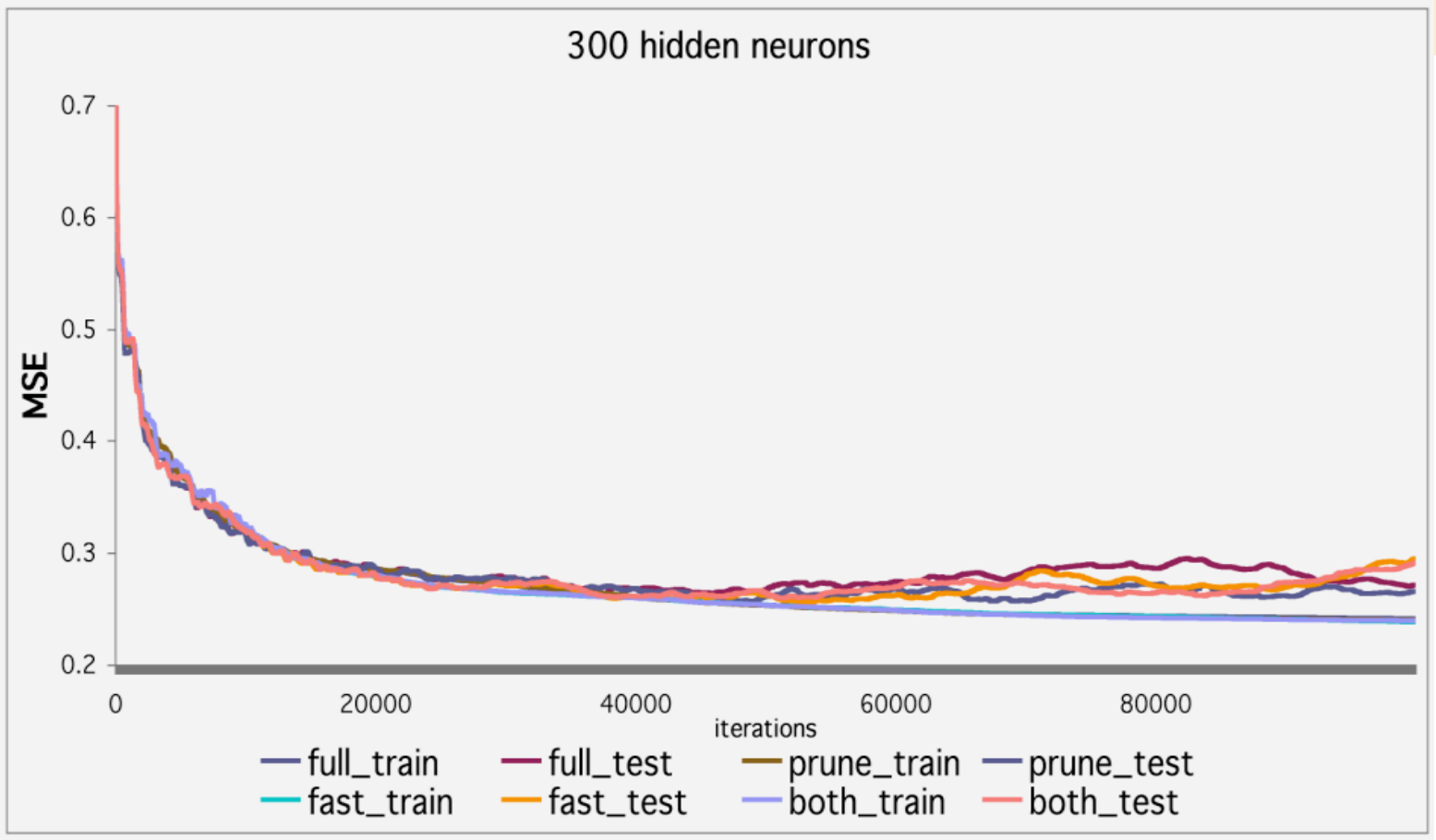


Fig. 7. 300 hidden neurons

Conclusions

- **The sooner the better:**
  - **Training Curve**: First, descend fast, then converge.
  - **Acceptable model**: When MSE starts to converge.
- **Lo-Shu data sampling can save time before convergence:**
  - Average time for at least one subset to reappear is less than two shifts away
  - Average time for at least two subsets to reappear is less than three shifts away

Fast Training MLP Networks with Lo-Shu Data Sampling

Q & A