

Ability to Generalize

- 1 Ability to Generalize
 - Overfitting
 - Early Stopping
 - Pruning
- 2 Alternate First Layers
 - Radial Basis Functions

- 1 Ability to Generalize
 - Overfitting
 - Early Stopping
 - Pruning
- 2 Alternate First Layers
 - Radial Basis Functions

Generalization

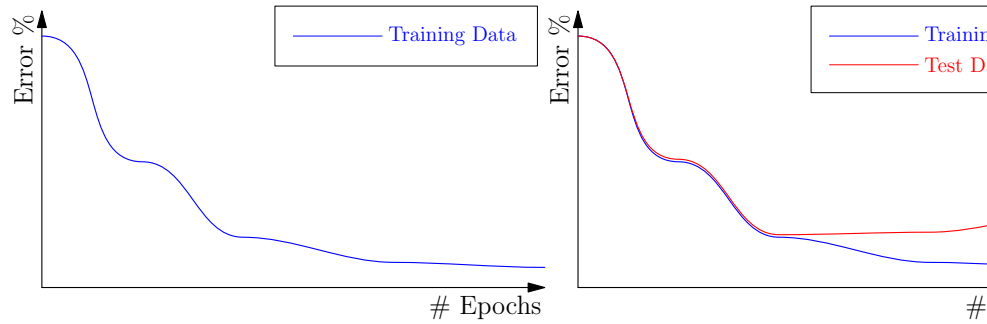
Ability to give the right answer for input not seen during training.

Two ways of measuring performance:

- Error frequency for training data
- Error frequency for novel data

Overfitting

Typical learning curve



When the model (the network) is powerful, the risk increases that it learns peculiarities specific for the training samples.

Overfitting

Risk increases with

- Few data points
- Many weights
- Unlimited training time

Overfitting

What can be done about it?

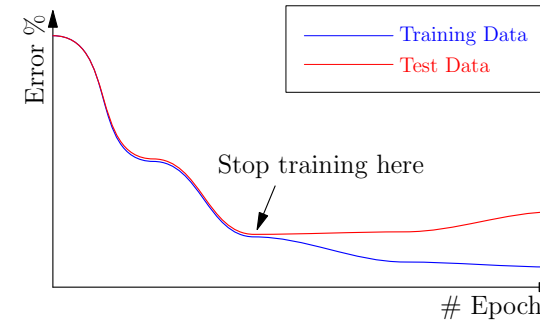
- Get more data (if possible)
- Restrict the number of hidden units
- Stop training prematurely ("Early Stopping")

Good News

Overfitting is *measurable*:

Compare performance on training and test data
"Cross Validation"

Overfitting



Early Stopping

Builds on the assumption that initial weights are too small

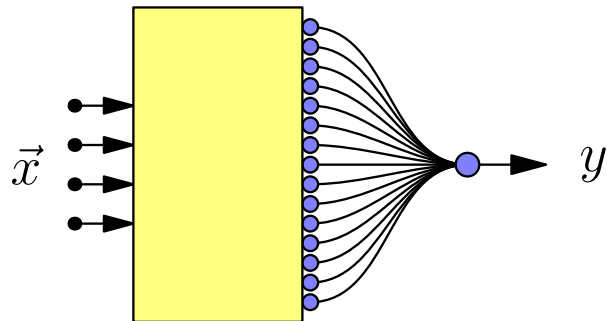
Pruning

Pruning

Techniques to reduce model complexity.

- Removal of unnecessary hidden units
- Removal of weak weights

Can be achieved by including a penalty term in the cost function



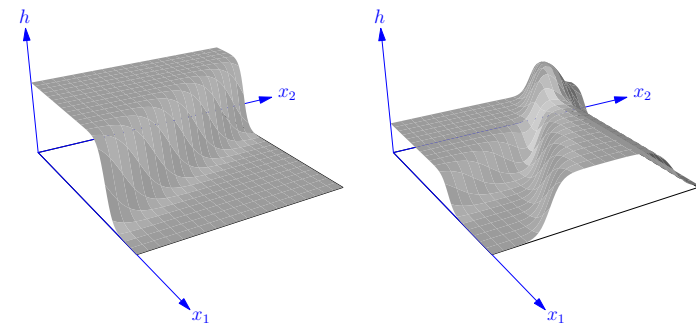
1 Ability to Generalize

- Overfitting
- Early Stopping
- Pruning

2 Alternate First Layers

- Radial Basis Functions

Weighted sum of Base Functions



Why not use better base functions?

Radial Basis Functions

- Local base functions
- Radially symmetric

