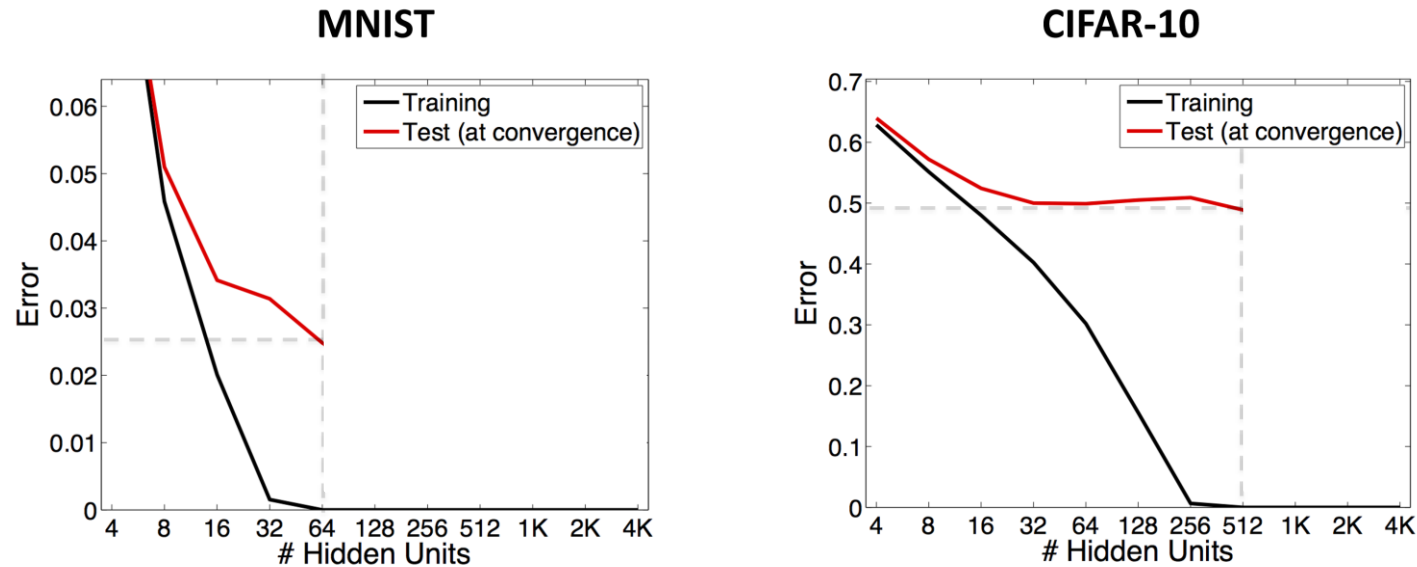


CPSC 440: Machine Learning

Implicit Regularization
Winter 2021

“Hidden” Regularization in Neural Networks

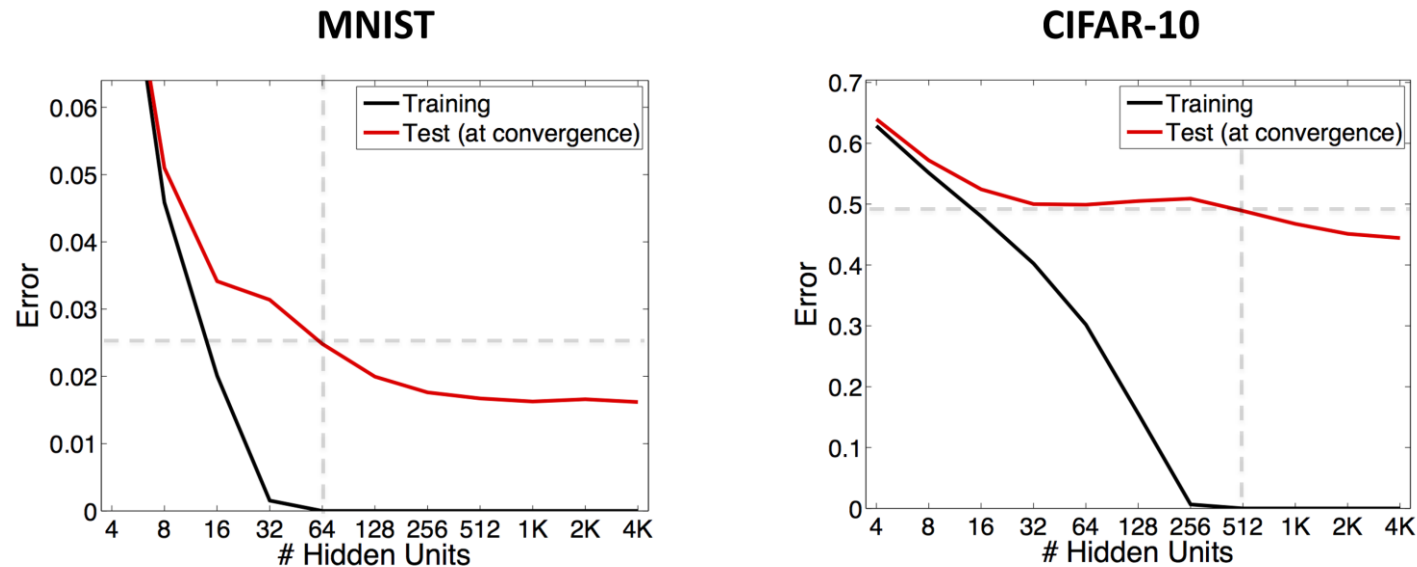
- Fitting **single-layer neural network with SGD and no regularization:**



- Training goes to 0 with enough units: **we're finding a global min.**
- What should happen to training and test error for larger #hidden?

“Hidden” Regularization in Neural Networks

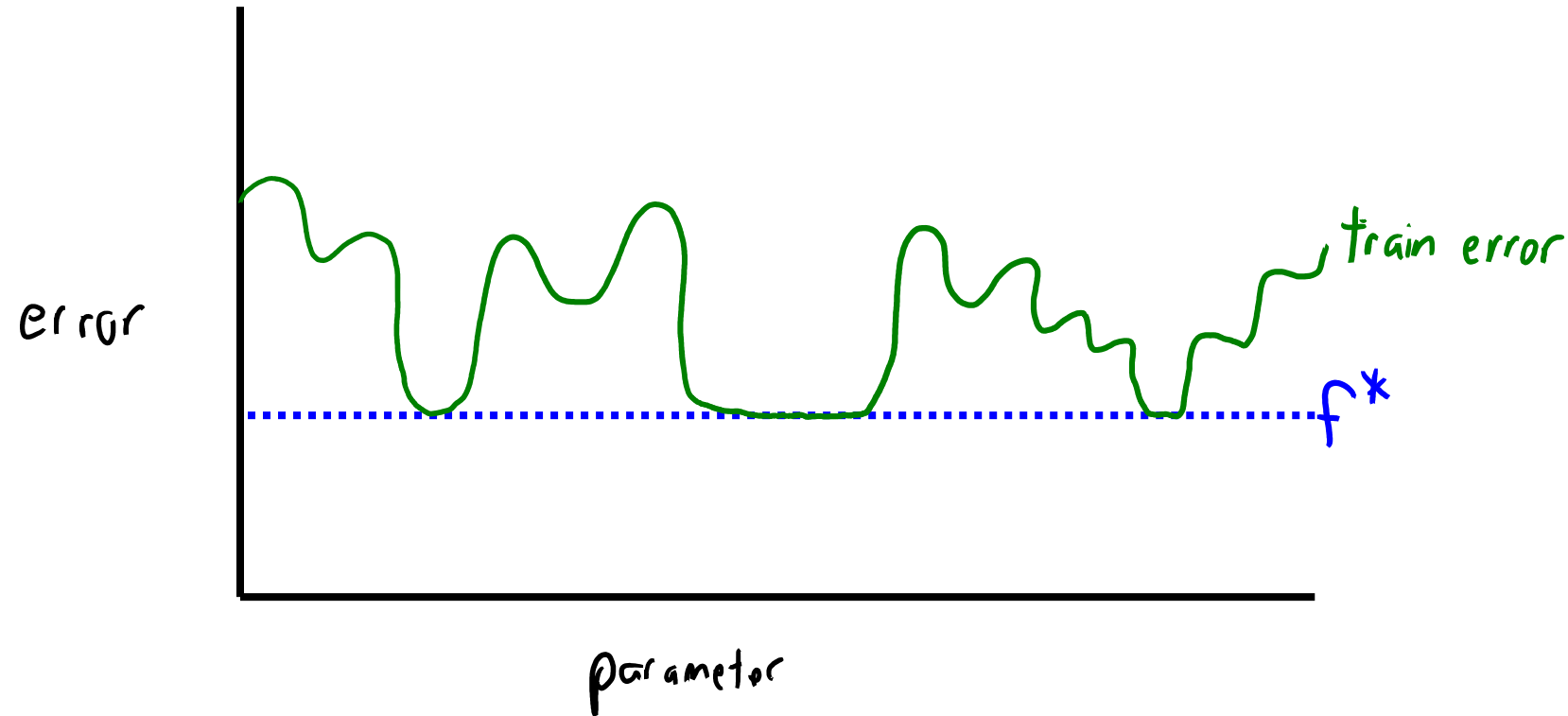
- Fitting single-layer neural network with SGD and no regularization:



- Test error continues to go down!?! Where is fundamental trade-off??
- There exist global mins with large #hidden units have test error = 1.
 - But among the global minima, SGD is somehow converging to “good” ones.

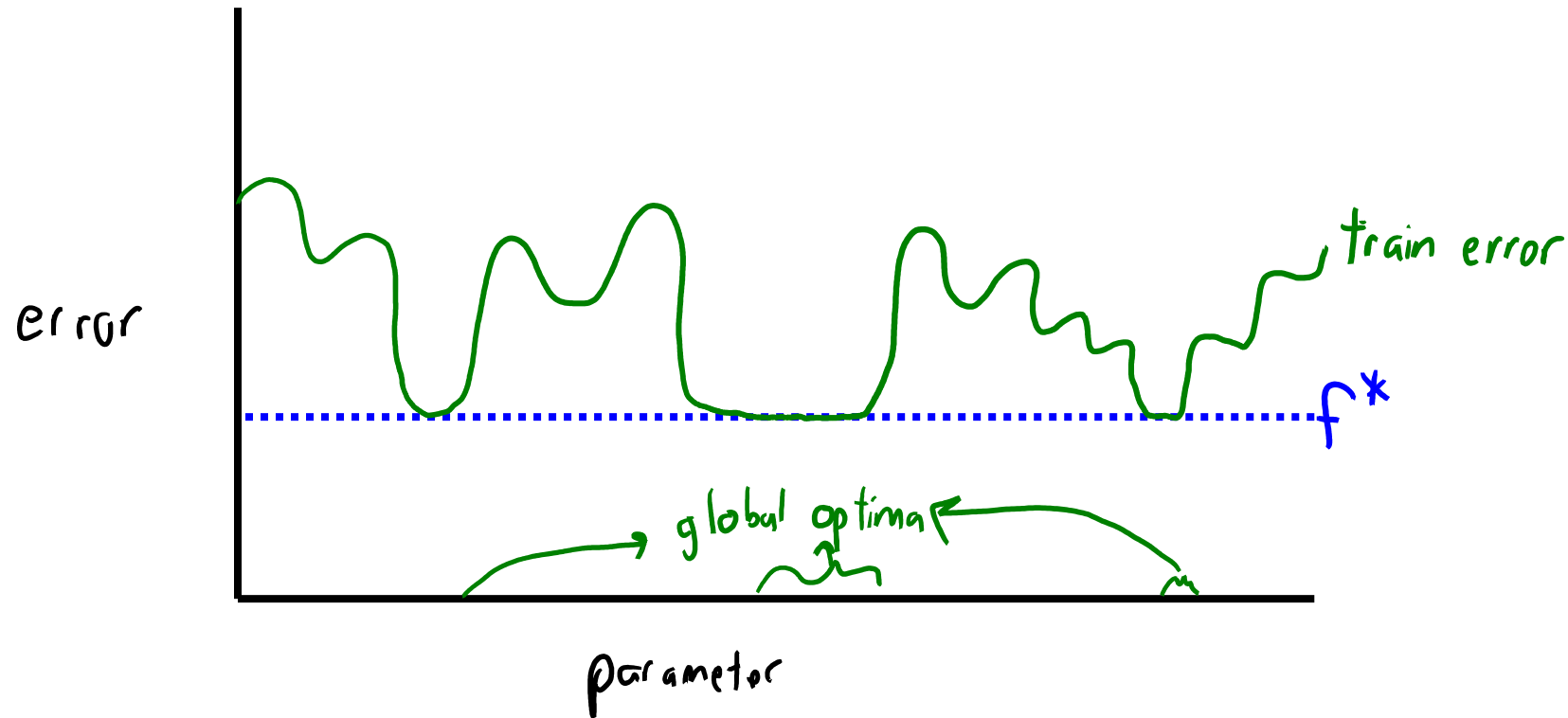
Multiple Global Minima?

- For standard objectives, there is a global min function value f^* :



Multiple Global Minima?

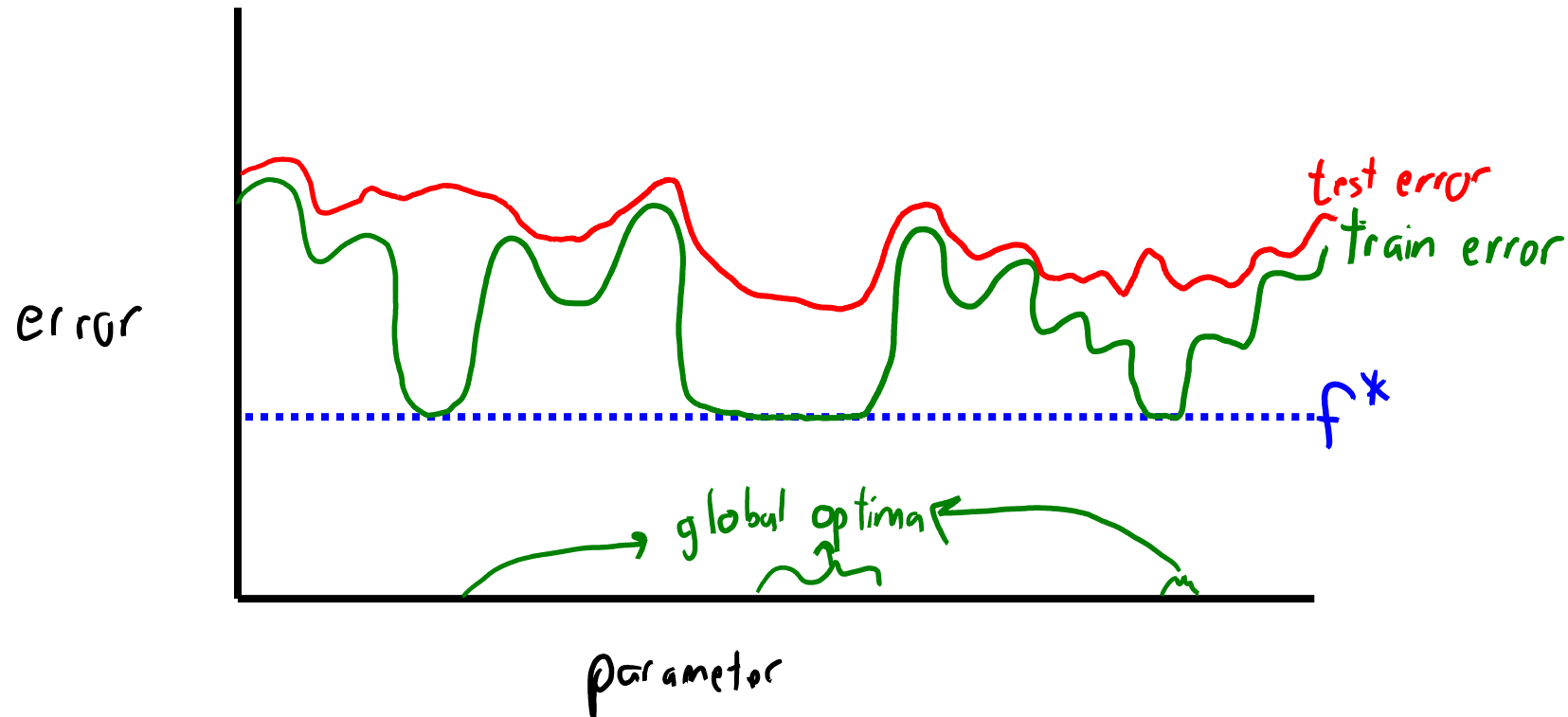
- For standard objectives, there is a global min function value f^* :



- But this may be achieved by many different parameter values.

Multiple Global Minima?

- For standard objectives, there is a global min function value f^* :



- But this may be achieved by many different parameter values.
 - These training error “global minima” may have very-different test errors.
 - Some of these global minima may be more “regularized” than others.

Implicit Regularization of SGD

- There is growing evidence that using SGD regularizes parameters.
 - We call this the “implicit regularization” of the optimization algorithm.