E9 205 Machine Learning for Signal Procesing

Considerations in Deep NN learning

15-11-2017

Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

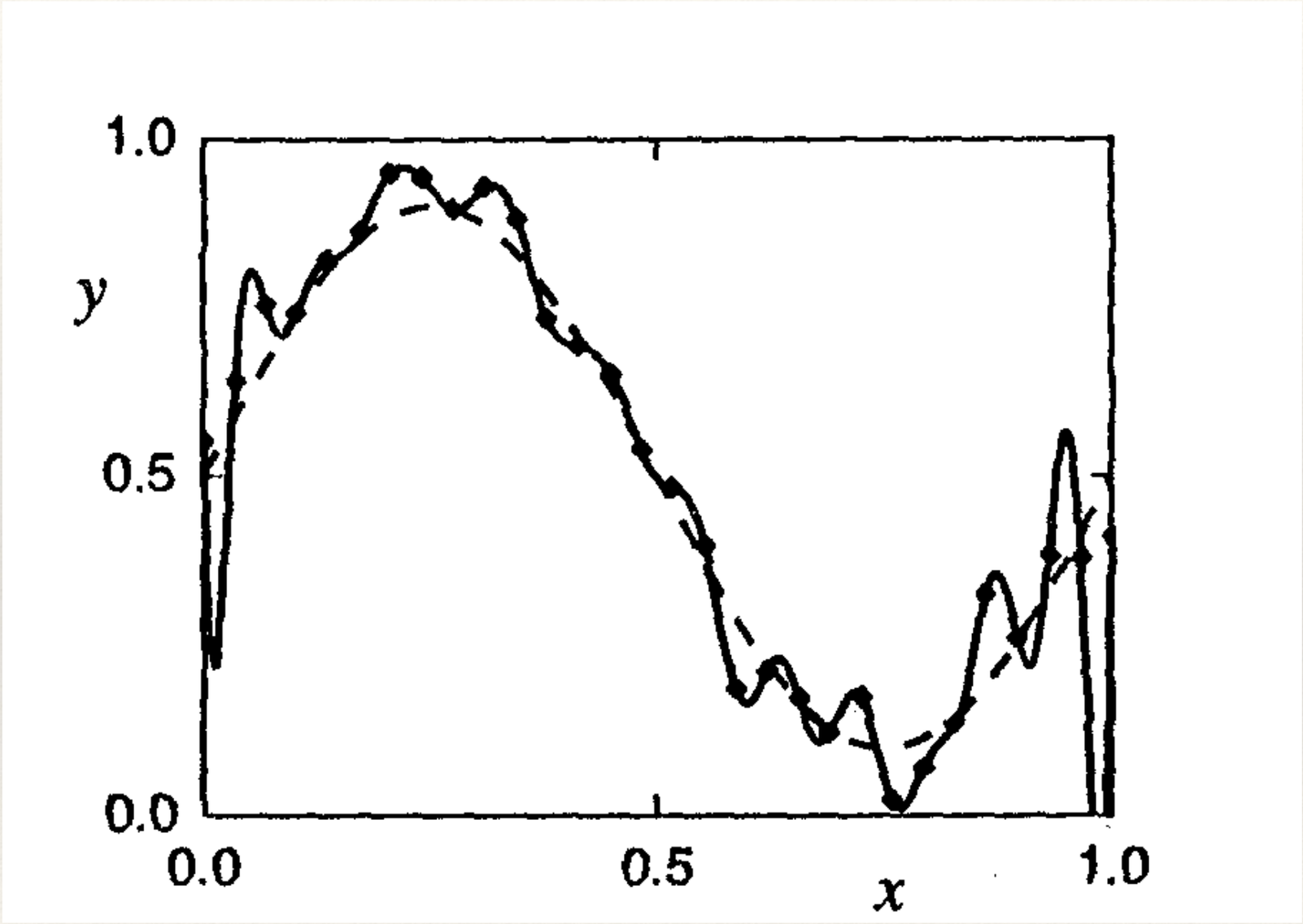
Bias Variance Trade Off and Overfitting

$$(\text{bias})^2 = \frac{1}{2}\int \{\mathcal{E}_D[y(\mathbf{x})] - \langle t|\mathbf{x}\rangle\}^2 p(\mathbf{x})\, d\mathbf{x}$$

$$\text{variance} = \frac{1}{2}\int \mathcal{E}_D[\{y(\mathbf{x}) - \mathcal{E}_D[y(\mathbf{x})]\}^2] p(\mathbf{x})\, d\mathbf{x}.$$

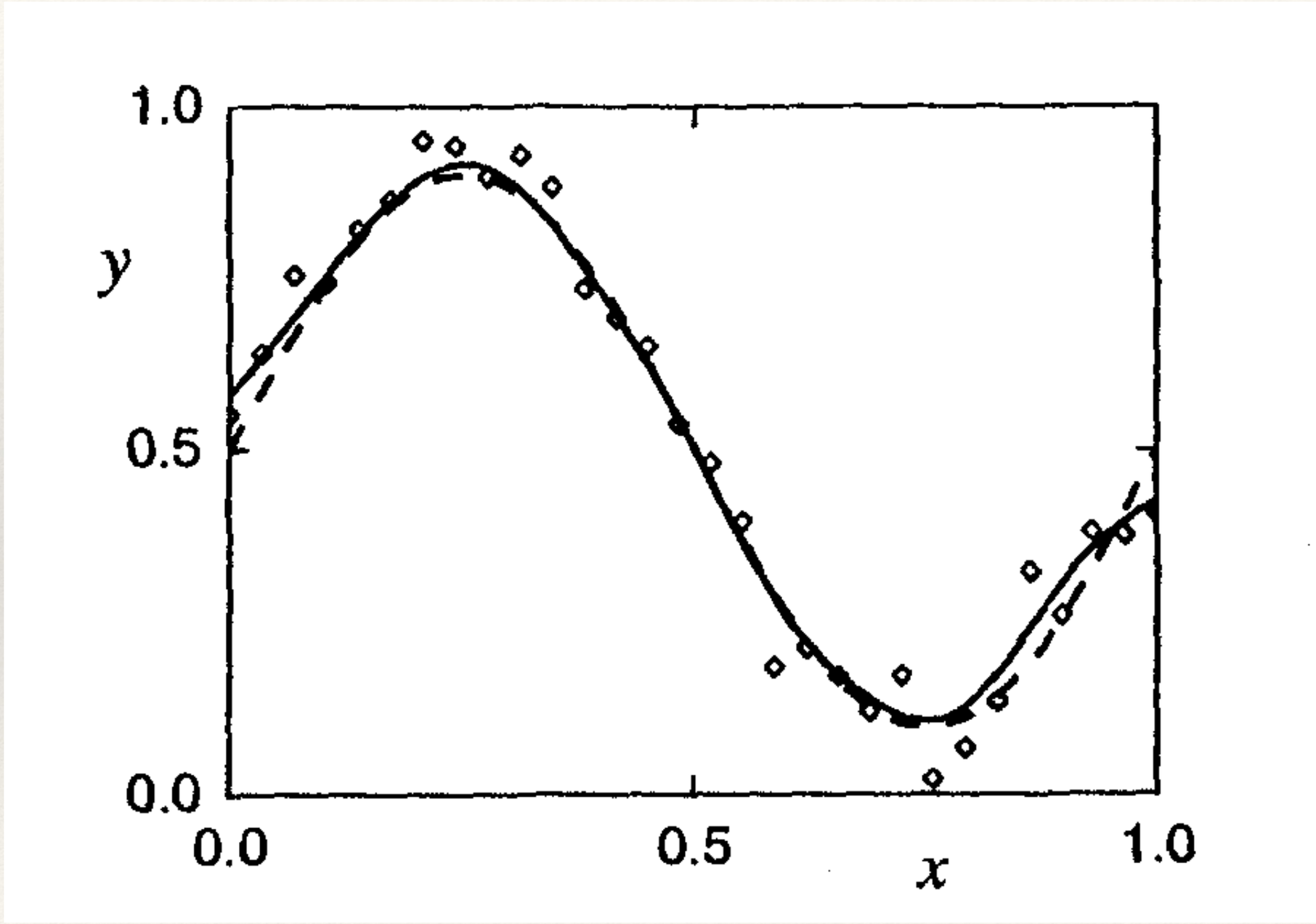
"Neural Networks", C. Bishop

Bias Variance Trade Off and Overfitting



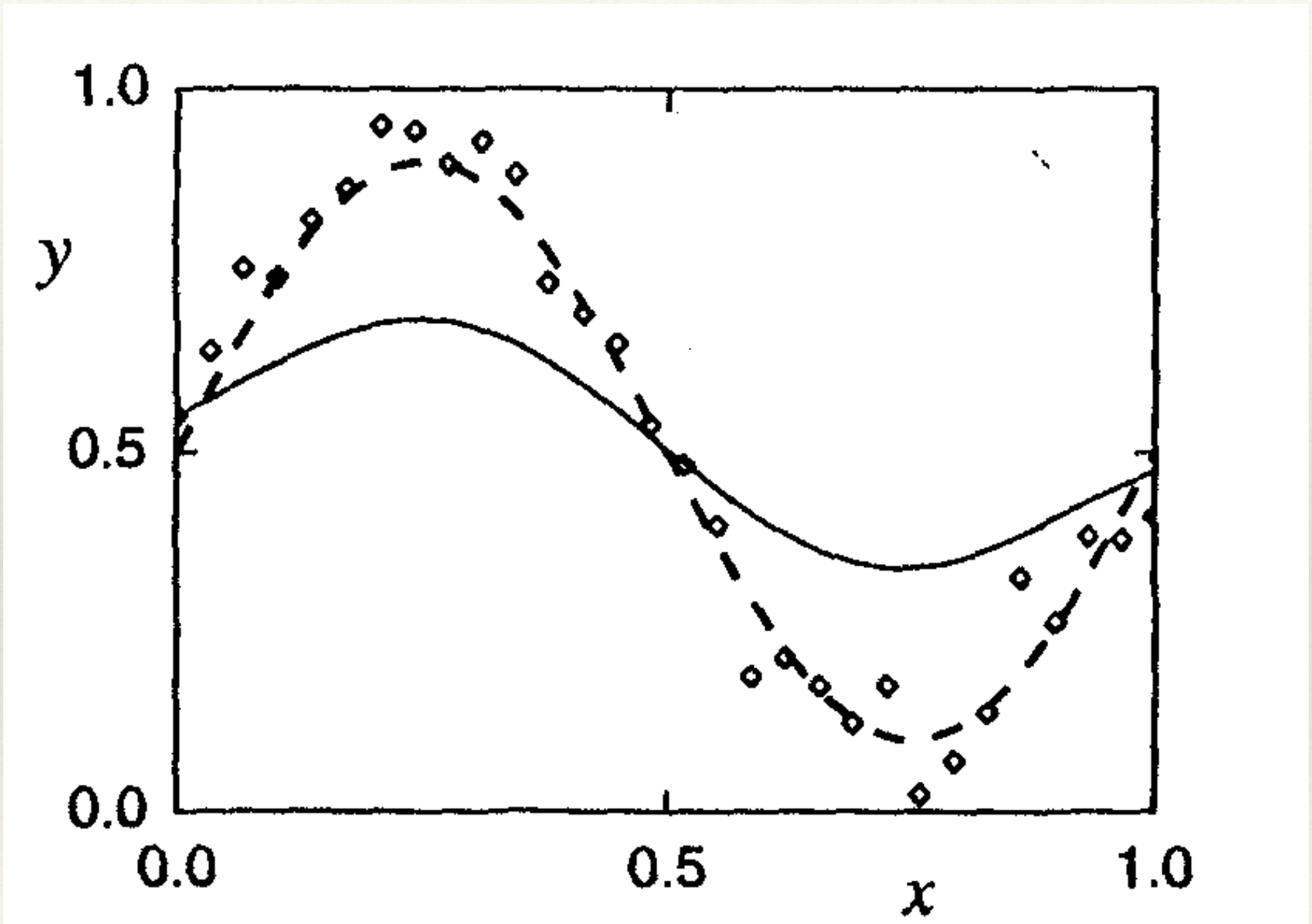
"Neural Networks", C. Bishop

Weight Decay Regularization



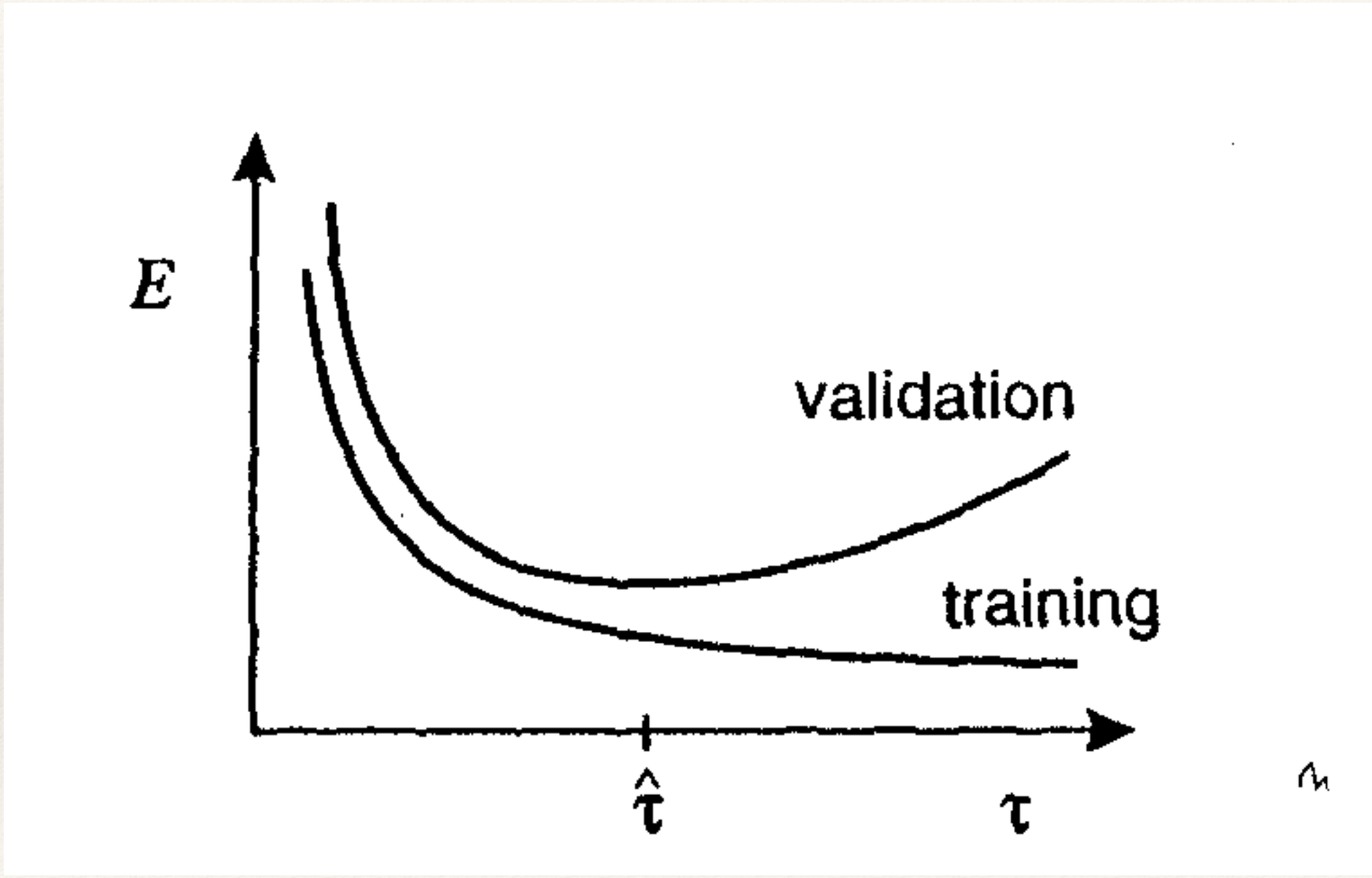
"Neural Networks", C. Bishop

Weight Decay Regularization



"Neural Networks", C. Bishop

Early Stopping

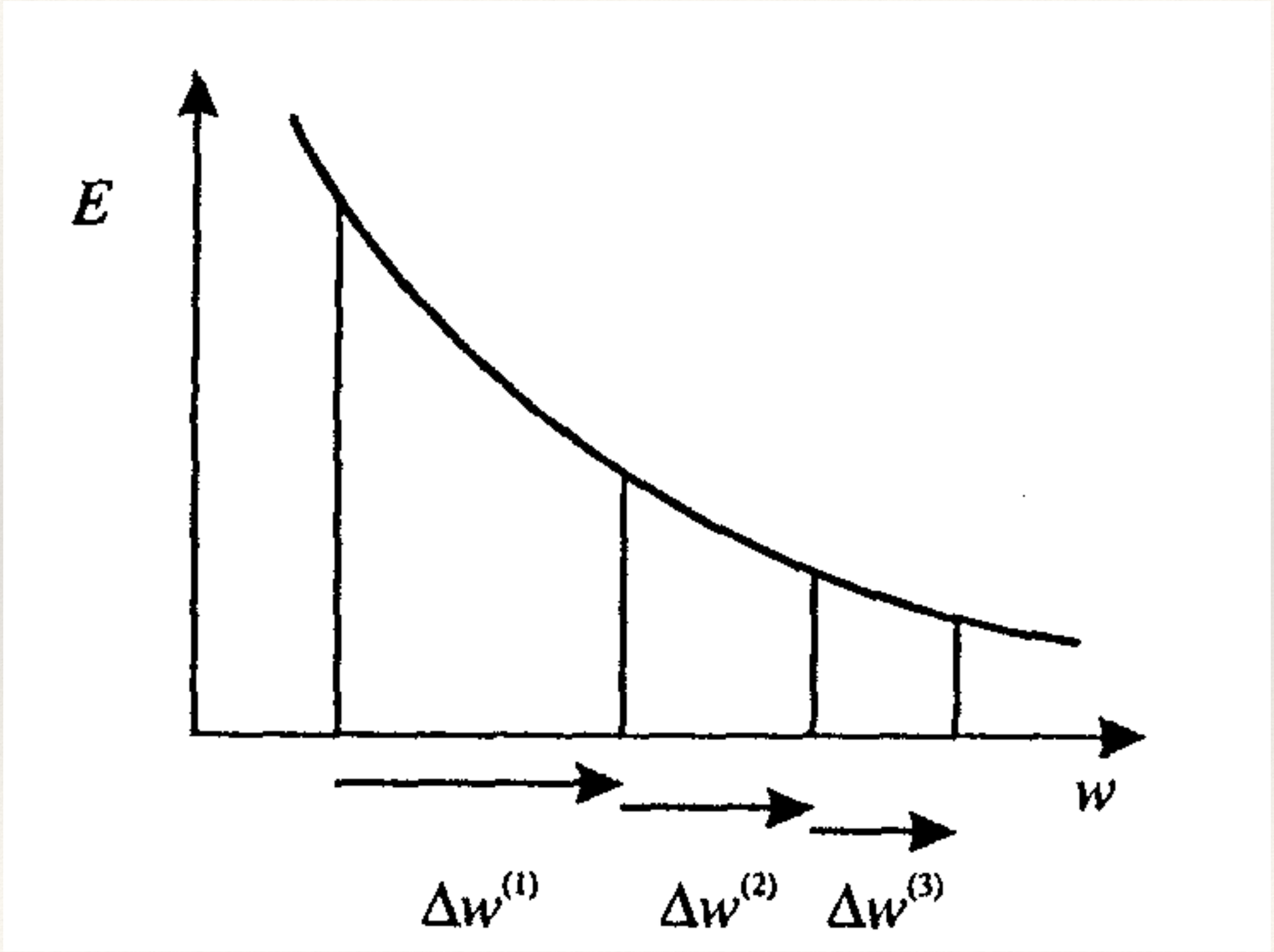


"Neural Networks", C. Bishop

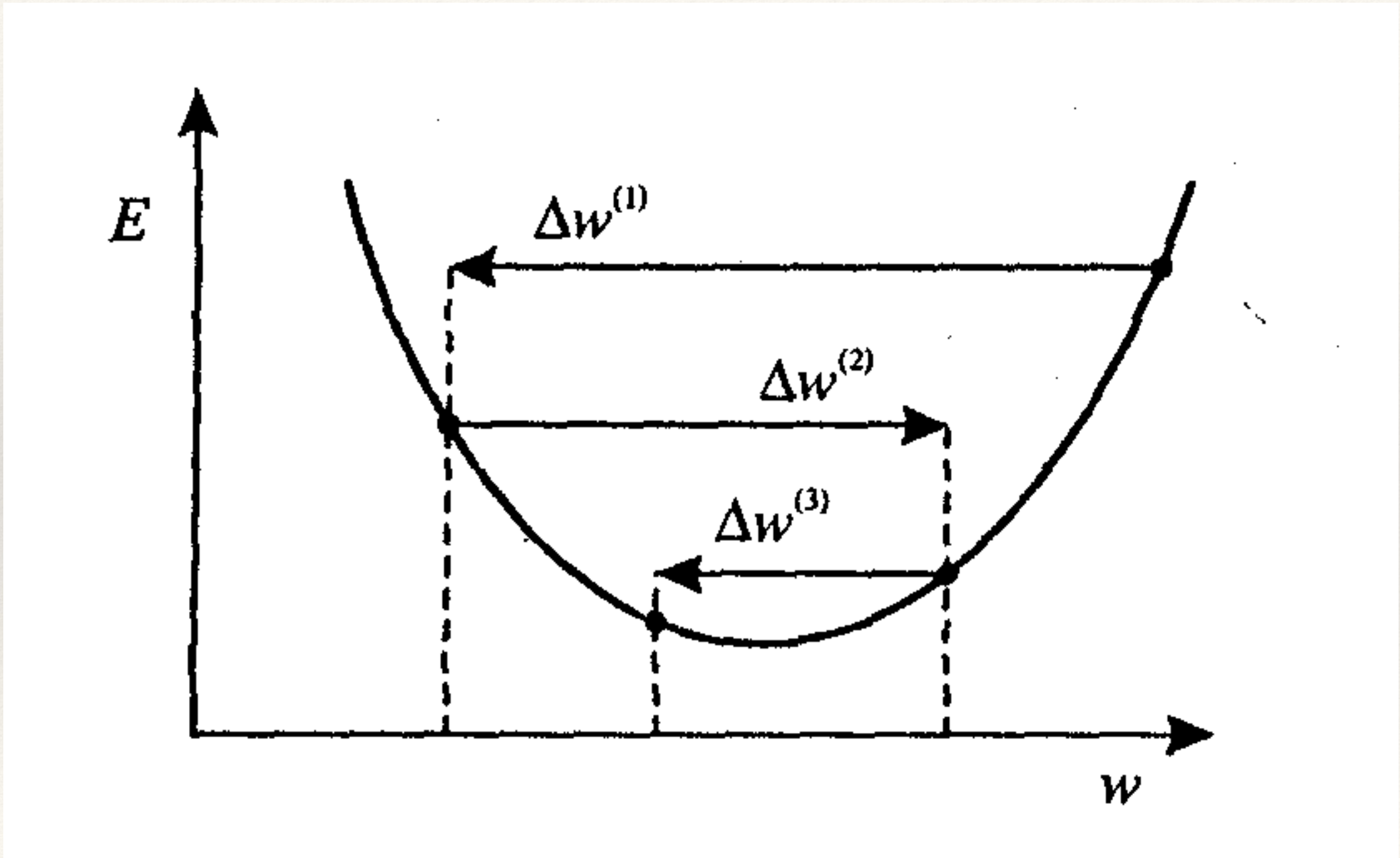
Attempts to Improve Learning

- Increasing the amount of training data
- Regularization (eg. weight decay)
- Early stopping using cross validation
- Training data with noise.
- Committees of Neural Networks

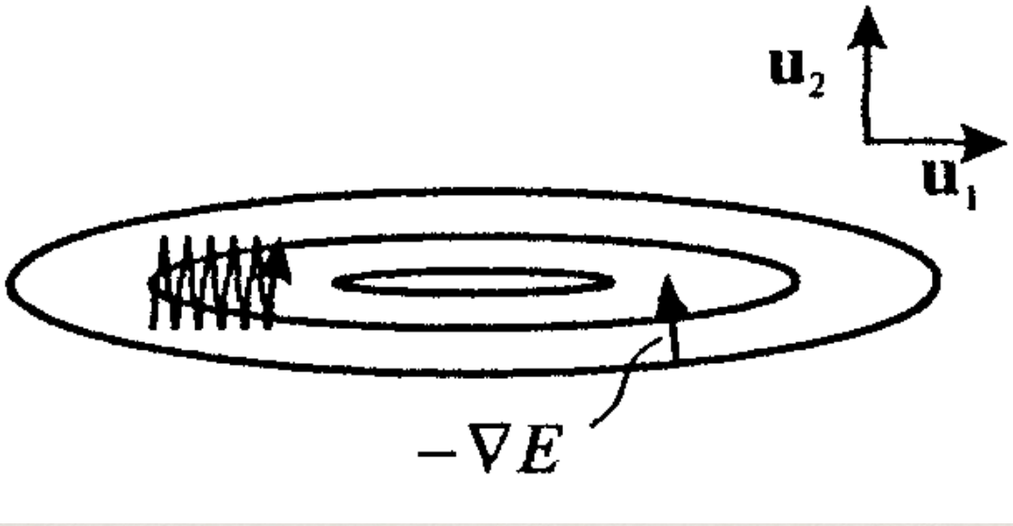
Momentum



Momentum

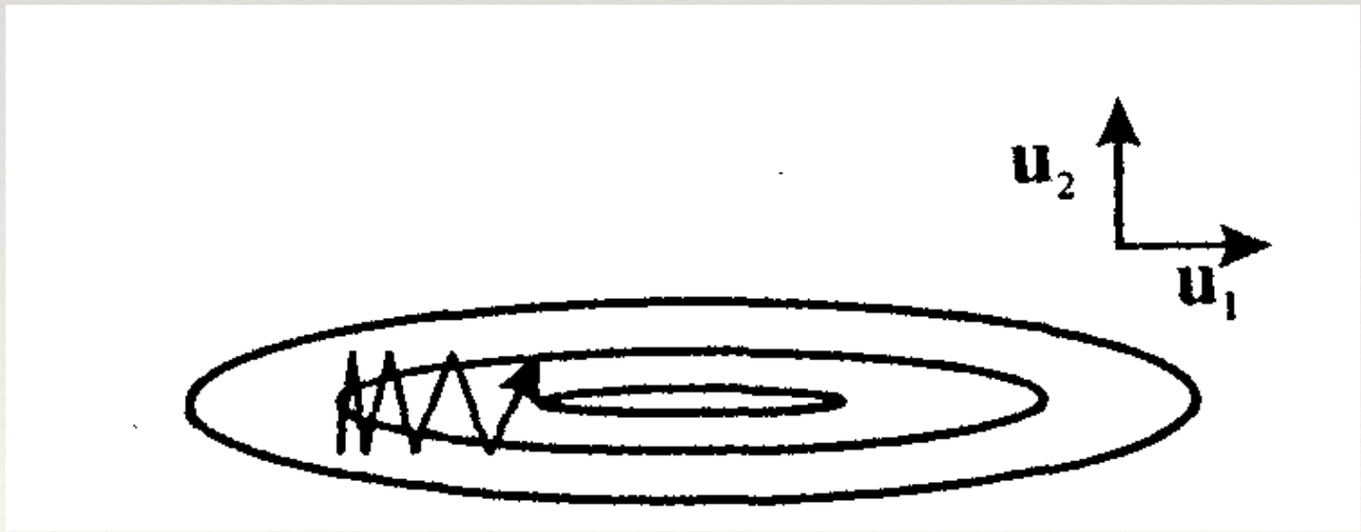


Momentum



Without Momentum

With Momentum

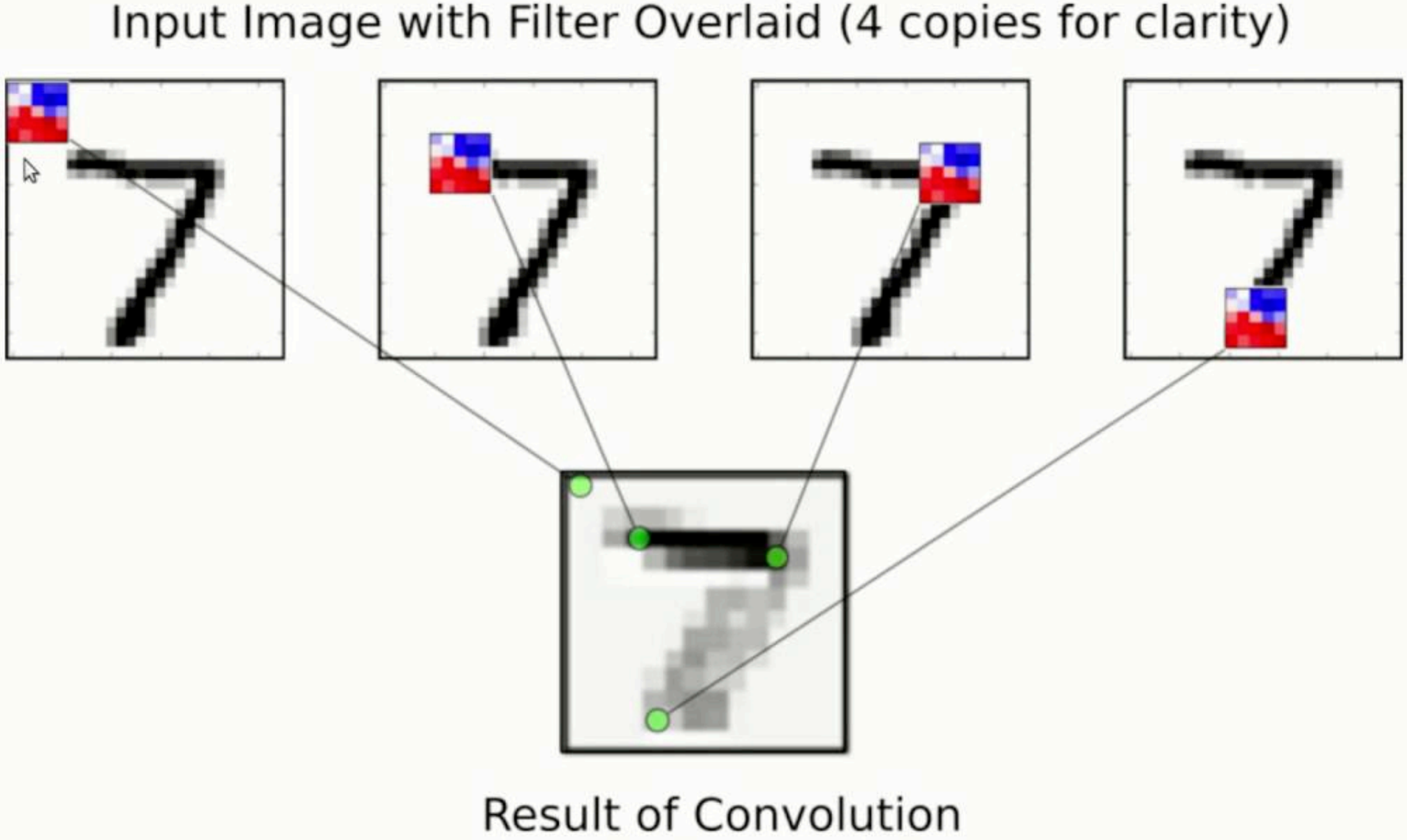


Convolution Operation (Images)

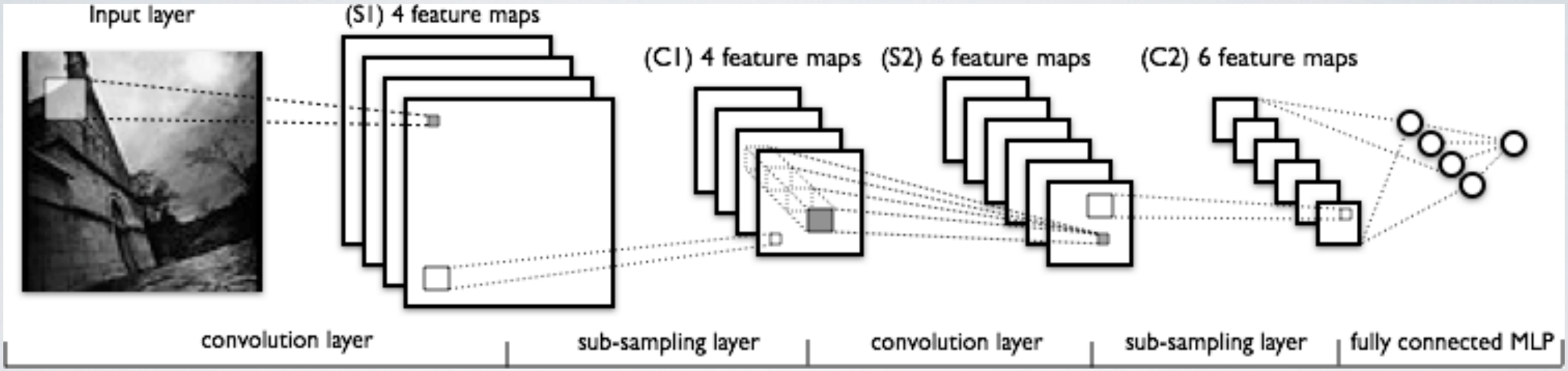
File Edit View Insert Cell Kernel Help Python 3

```
Image('images/02_convolution.png')
```

Out[2]:



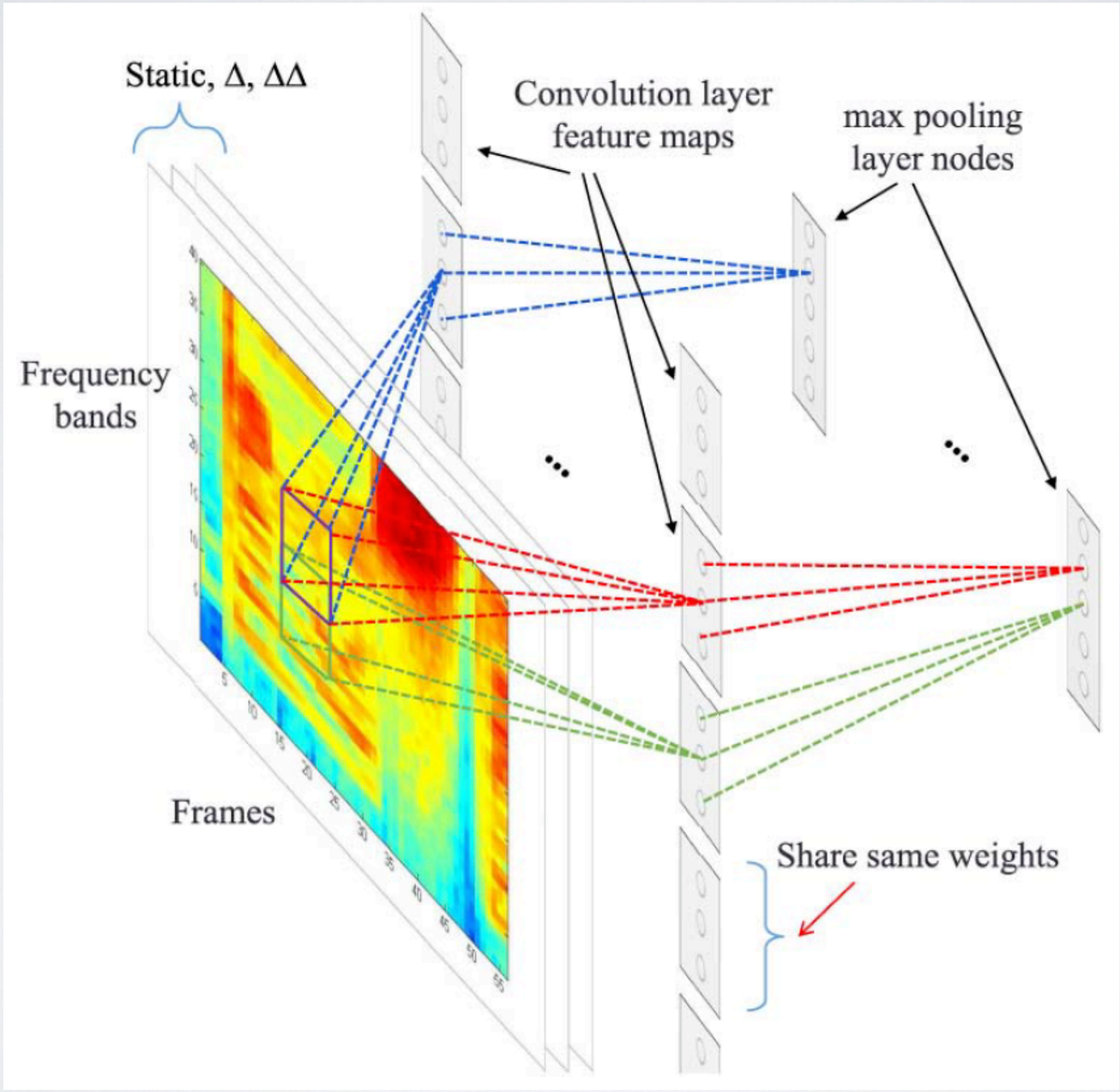
Convolutional Networks



- Multiple levels of filtering and subsampling operations.
- Feature maps are generated at every layer.

"deeplearning.net"

CNNs for Speech



"Speech Recognition Wiki"