# *E9 205 Machine Learning for Signal Processing*

**Deep Learning**

23-11-2016

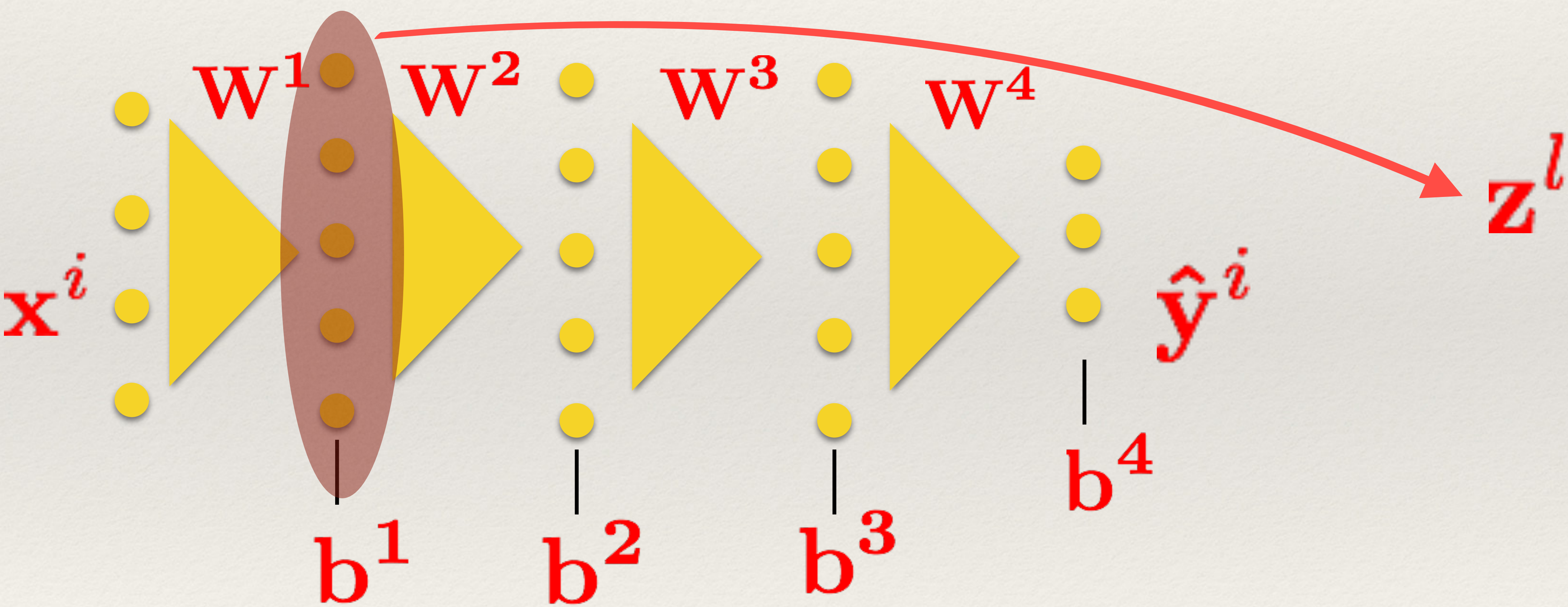
Established 1911

INDIAN INSTITUTE OF SCIENCE

भारतीय विज्ञान संस्थान

# Deep Feed Forward Networks

Neural networks with multiple hidden layers - Deep networks



# Back Propagation Algorithm

**Variables**

$\mathbf{x}^i$ Input of DNN

$\hat{\mathbf{y}}^i$ Output of DNN

$\mathbf{y}^i$ Desired Output

$\mathbf{W}_t^l$ $\mathbf{b}_t^l$ Weights layer $l$ iteration $t$

$m_b$ Mini-batch size

$\mathbf{z}_t^l$ Hidden layer output at $l$ iteration $t$

$\mathbf{h}_t^l$ Hidden layer output at $l$ iteration $t$

$J$ Cost Function

$g(.)$ Hidden Activations

**Gradient Descent Algorithm**

$$\mathbf{W}_{t+1}^l = \mathbf{W}_t^l - \epsilon \Delta \mathbf{W}_t^l$$

$$\mathbf{b}_{t+1}^l = \mathbf{b}_t^l - \epsilon \Delta \mathbf{b}_t^l$$

Define $\mathbf{e}_t^l = \dfrac{\partial J}{\partial \mathbf{h}_t^l}$

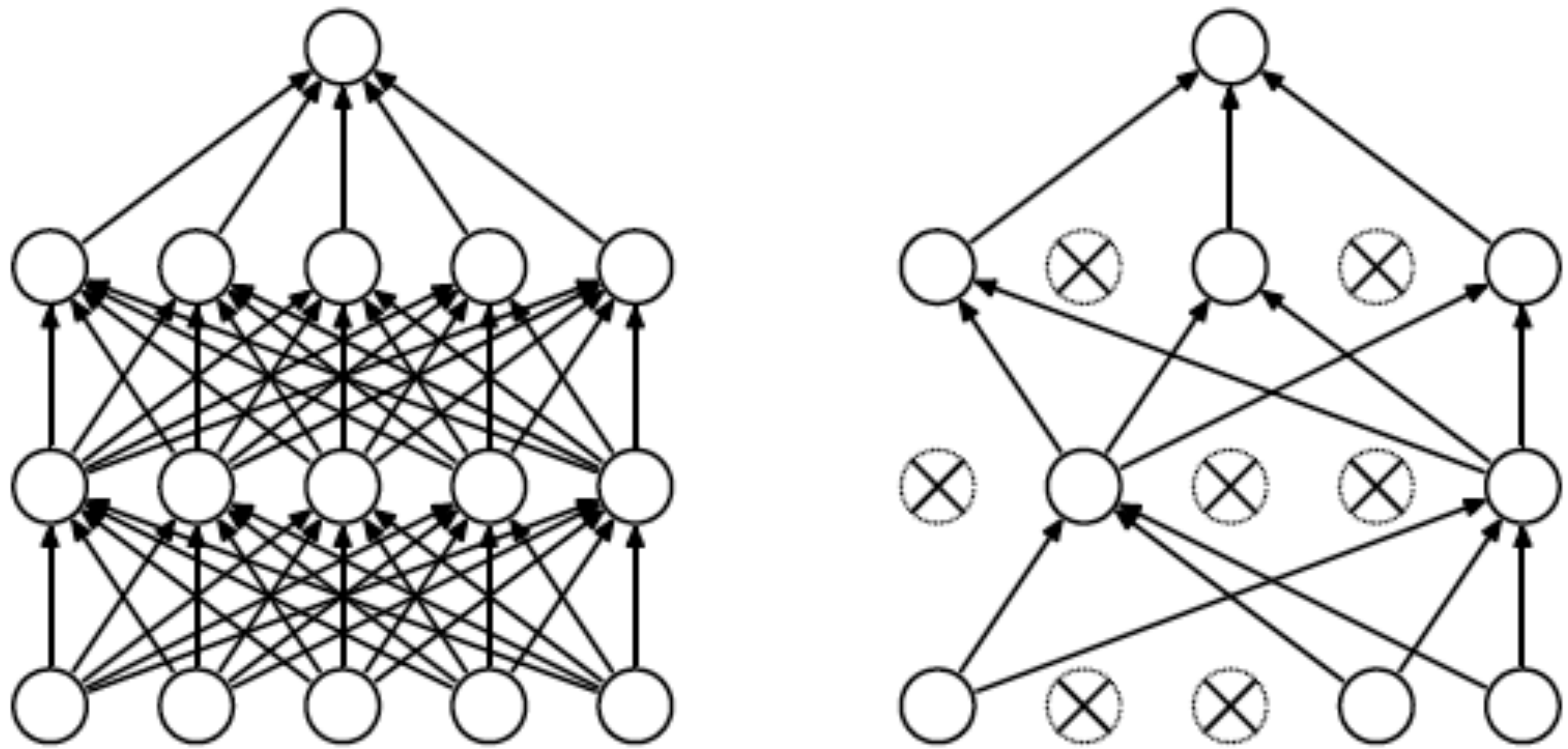
Then, $\mathbf{e}_t^{l-1} = (\mathbf{W}_t^l)^T \{g'(\mathbf{z}_t^l).\mathbf{e}_t^l\}$

$$\Delta_{\mathbf{W}_t^l} J = \{g'(\mathbf{z}_t^l).\mathbf{e}_t^l\}(\mathbf{h}_t^{l-1})^T$$

$$\Delta_{\mathbf{b}_t^l} J = \{g'(\mathbf{z}_t^l).\mathbf{e}_t^l\}$$

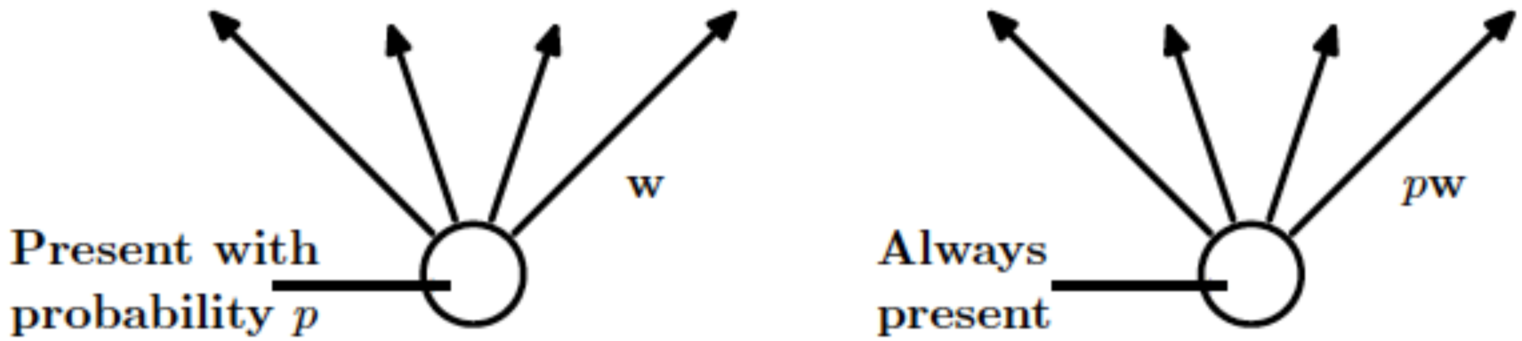
- Element wise product

# Regularization - Dropout



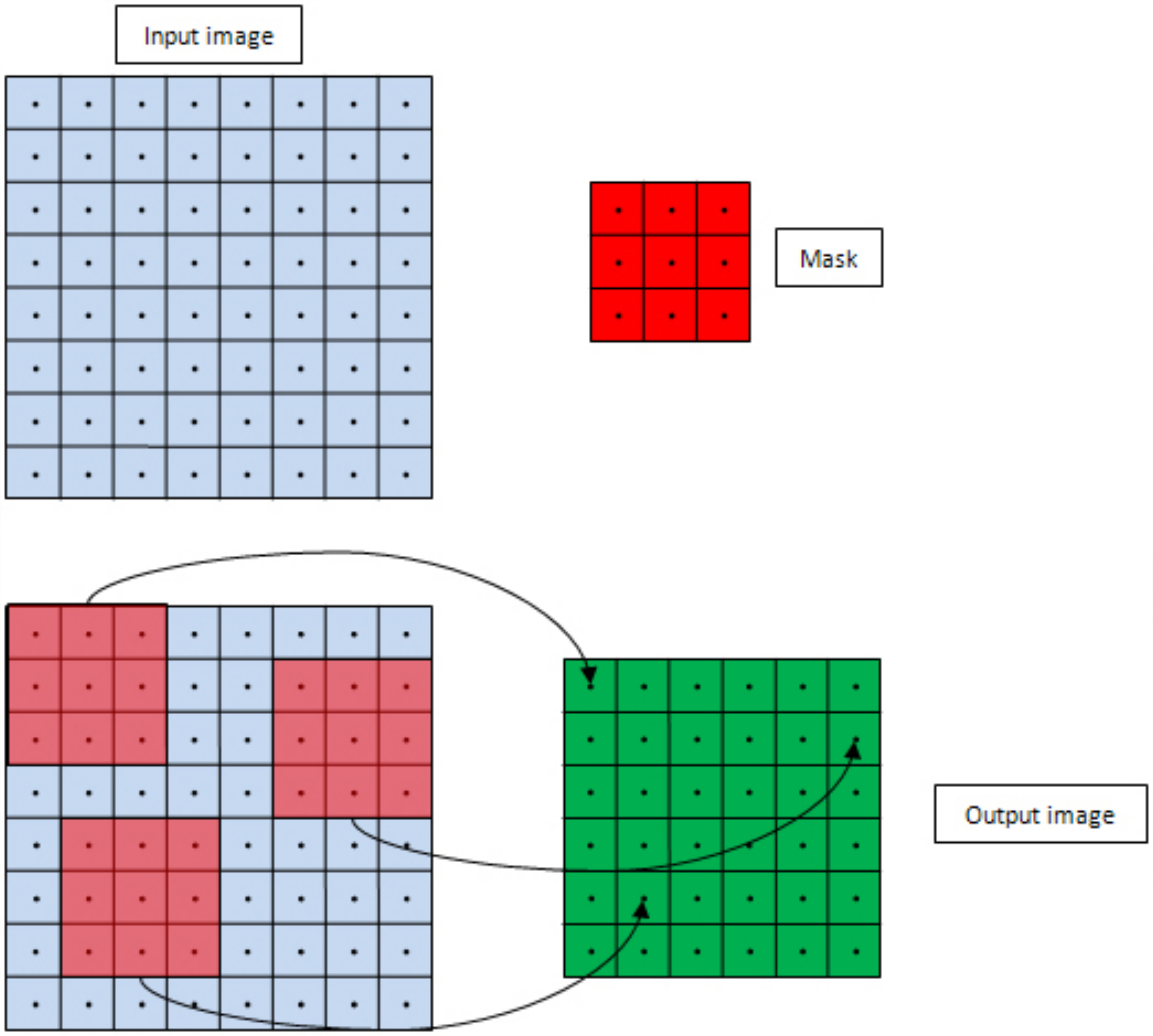
(a) Standard Neural Net (b) After applying dropout.

[Srivatsava, 2014]




(a) At training time (b) At test time

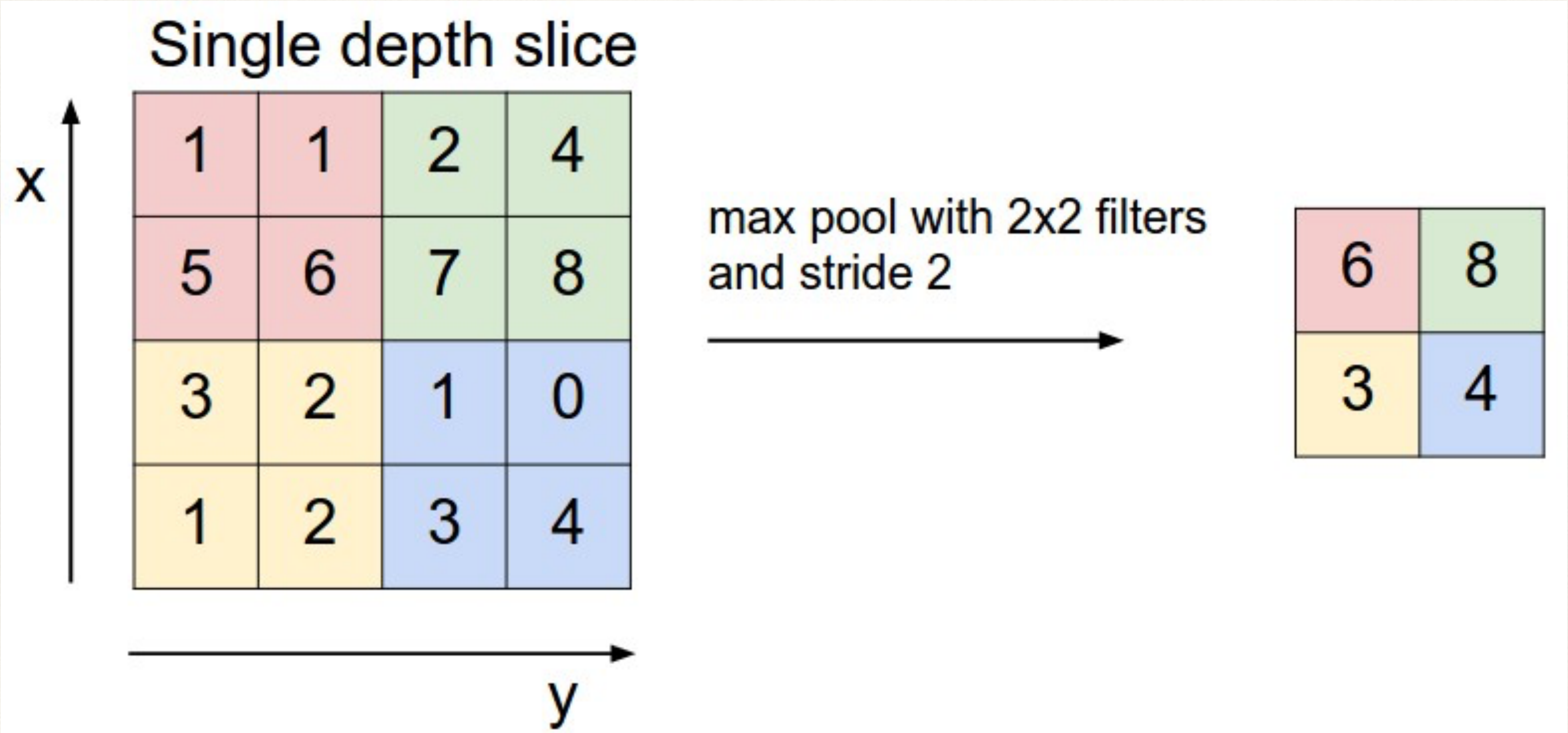
# Convolutional Network



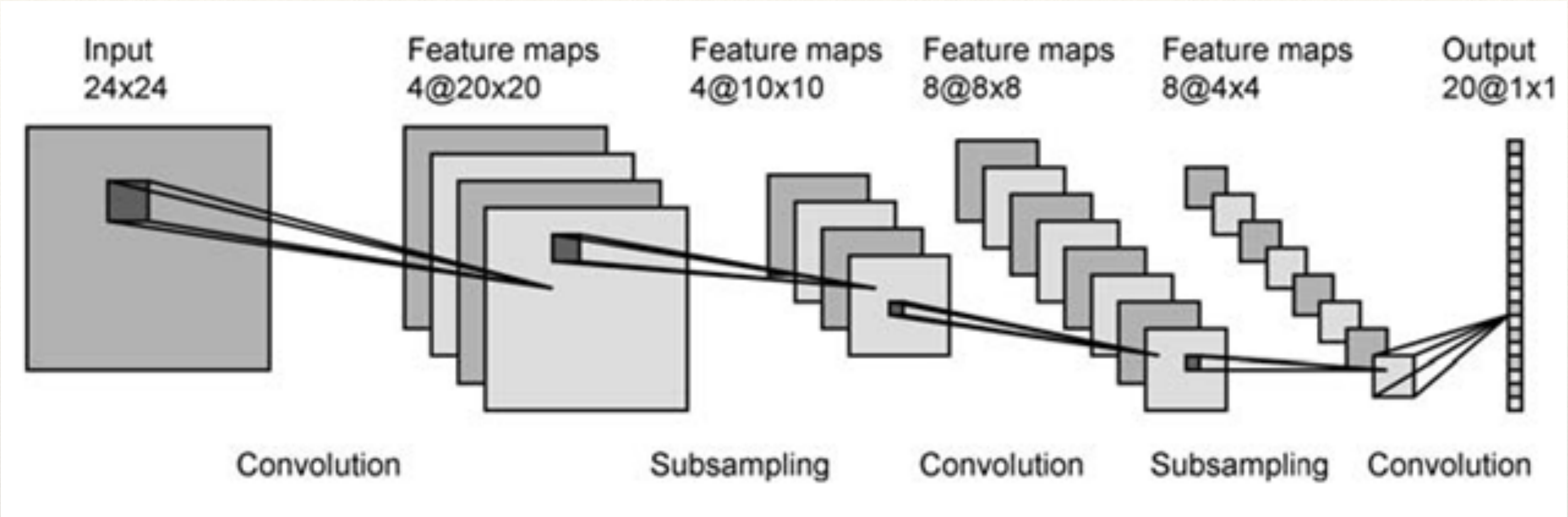

Weight sharing

# Convolutional Neural Network

## Max Pooling

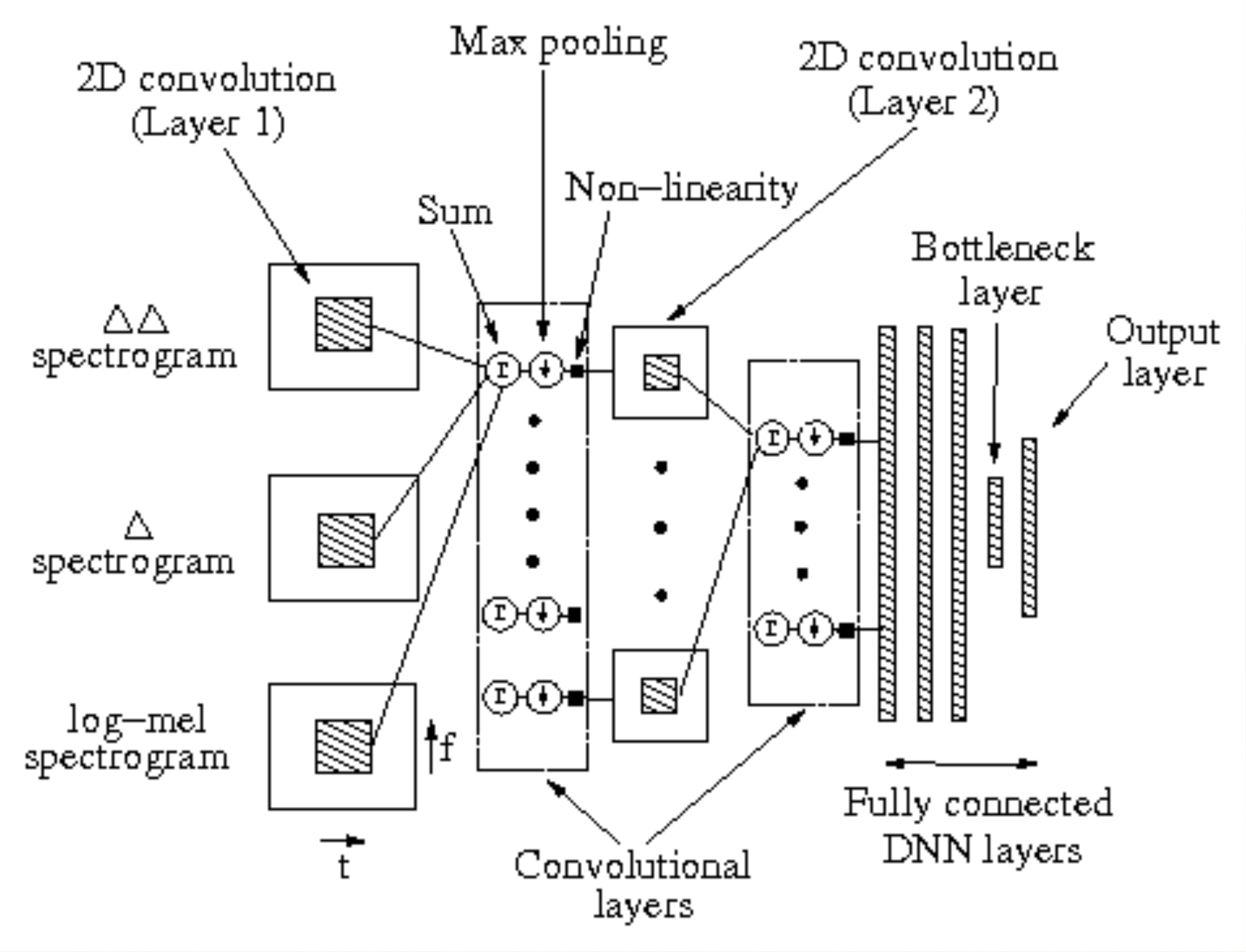


# Convolutional Neural Network

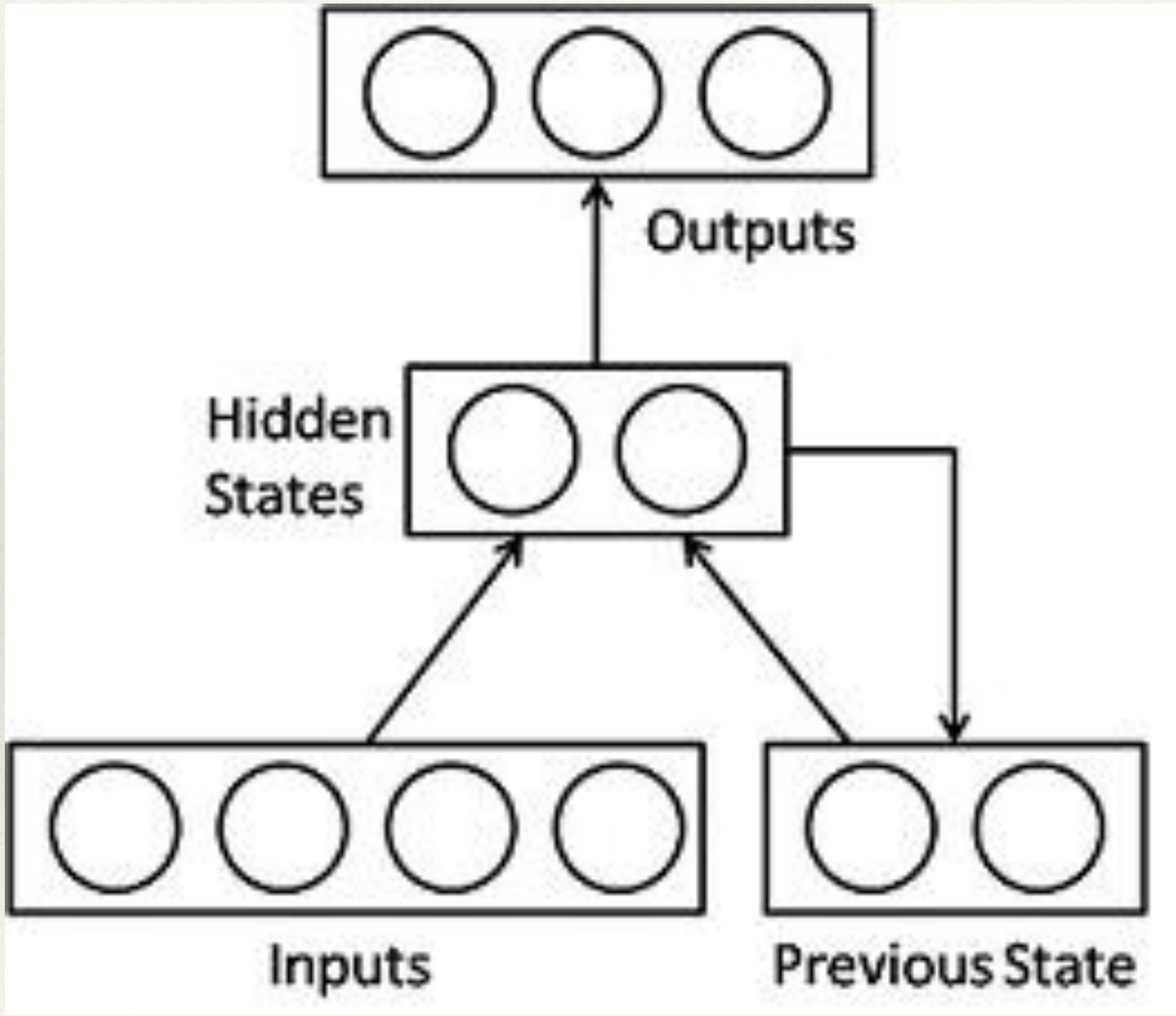



- Successive layers of convolutions and subsampling operation [LeCun, 98].
- Non-linearities (ReLu or Sigmoid) are typically applied after each subsampling stage.

# CNNs in Speech Processing



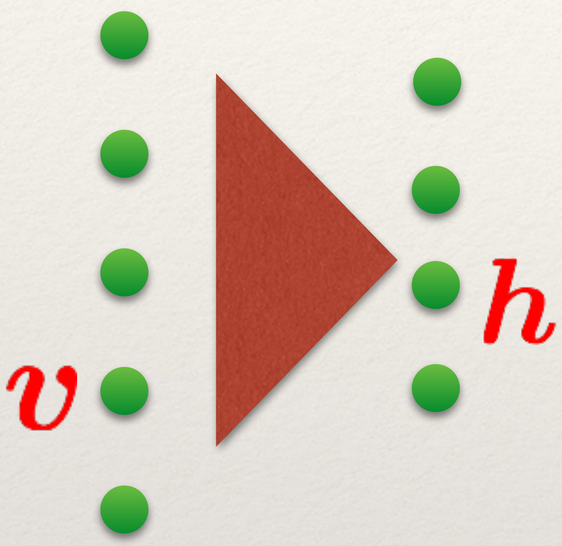
# Recurrent Networks




Feedback

# Restricted Boltzmann Machines

Initializing large networks with deep belief networks (DBN)

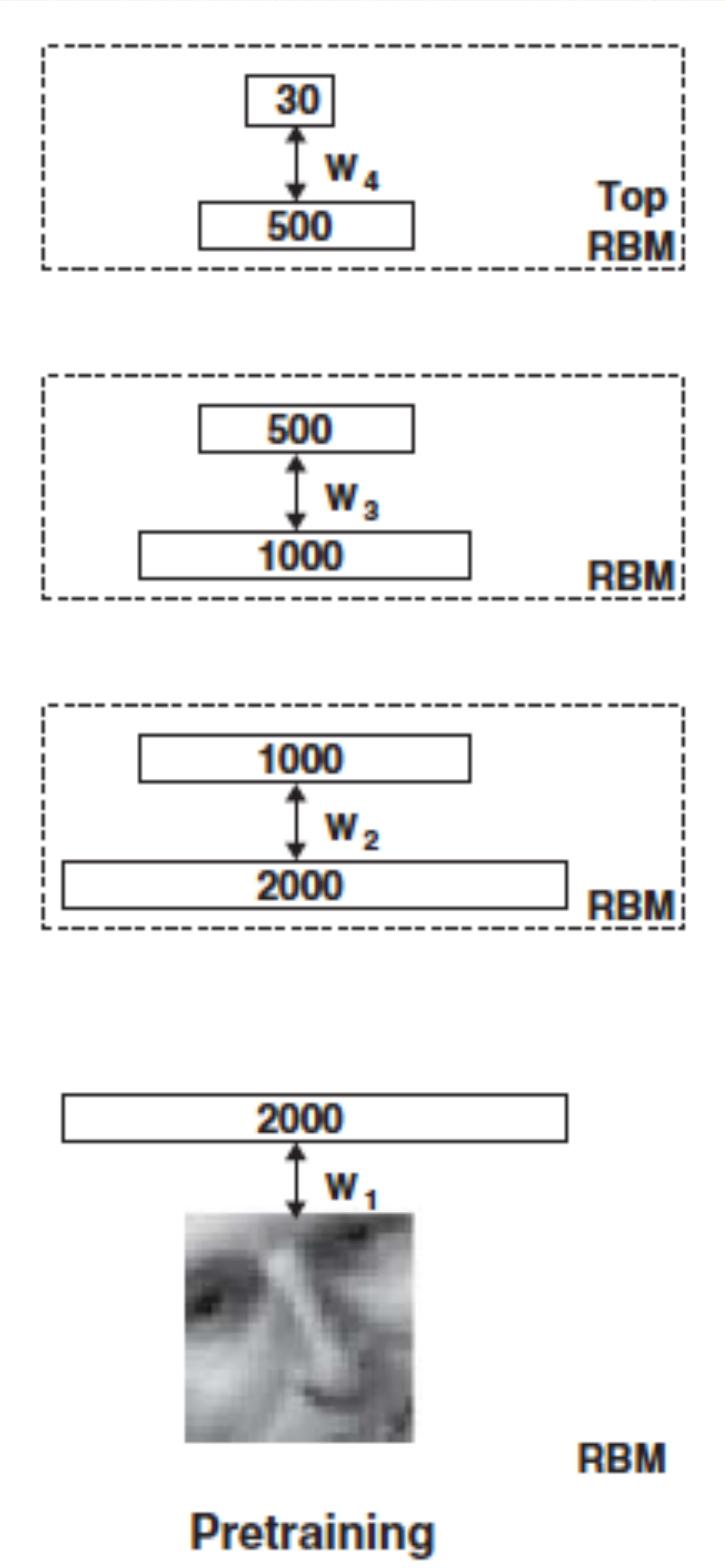



- Gaussian Bernoulli RBM - Gaussian visible layer and Bernoulli hidden layer.
- Define an energy function

$$E(\mathbf{v},\mathbf{h}) = 0.5(\mathbf{v}-\mathbf{a})^T(\mathbf{v}-\mathbf{a}) - \mathbf{b}^T\mathbf{h} - \mathbf{h}^T\mathbf{W}\mathbf{v}$$

- Define joint probability density $P(\mathbf{v},\mathbf{h}) = \dfrac{e^{-E(\mathbf{v},\mathbf{h})}}{Z}$
- Estimate the parameters $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{W}$ by maximizing the joint density

# Restricted Boltzmann Machines



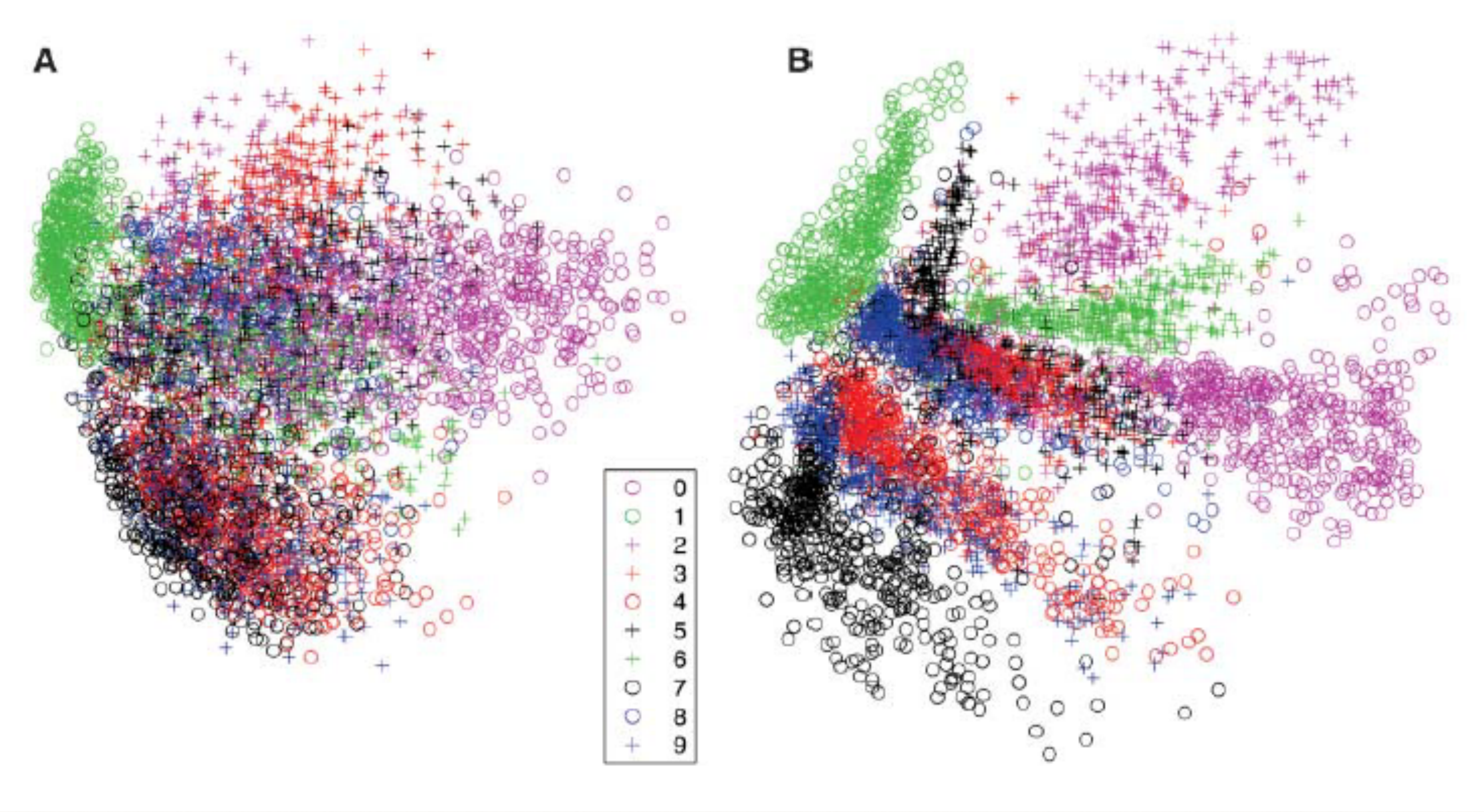

It can be shown that probability of hidden unit being is the sigmoid function

$$P(h_i = 1|\mathbf{v}) = \frac{1}{1 + e^{-(w_i^T \mathbf{v} + b_i)}}$$

RBM layers can be stacked one above the other in a hierarchical fashion.

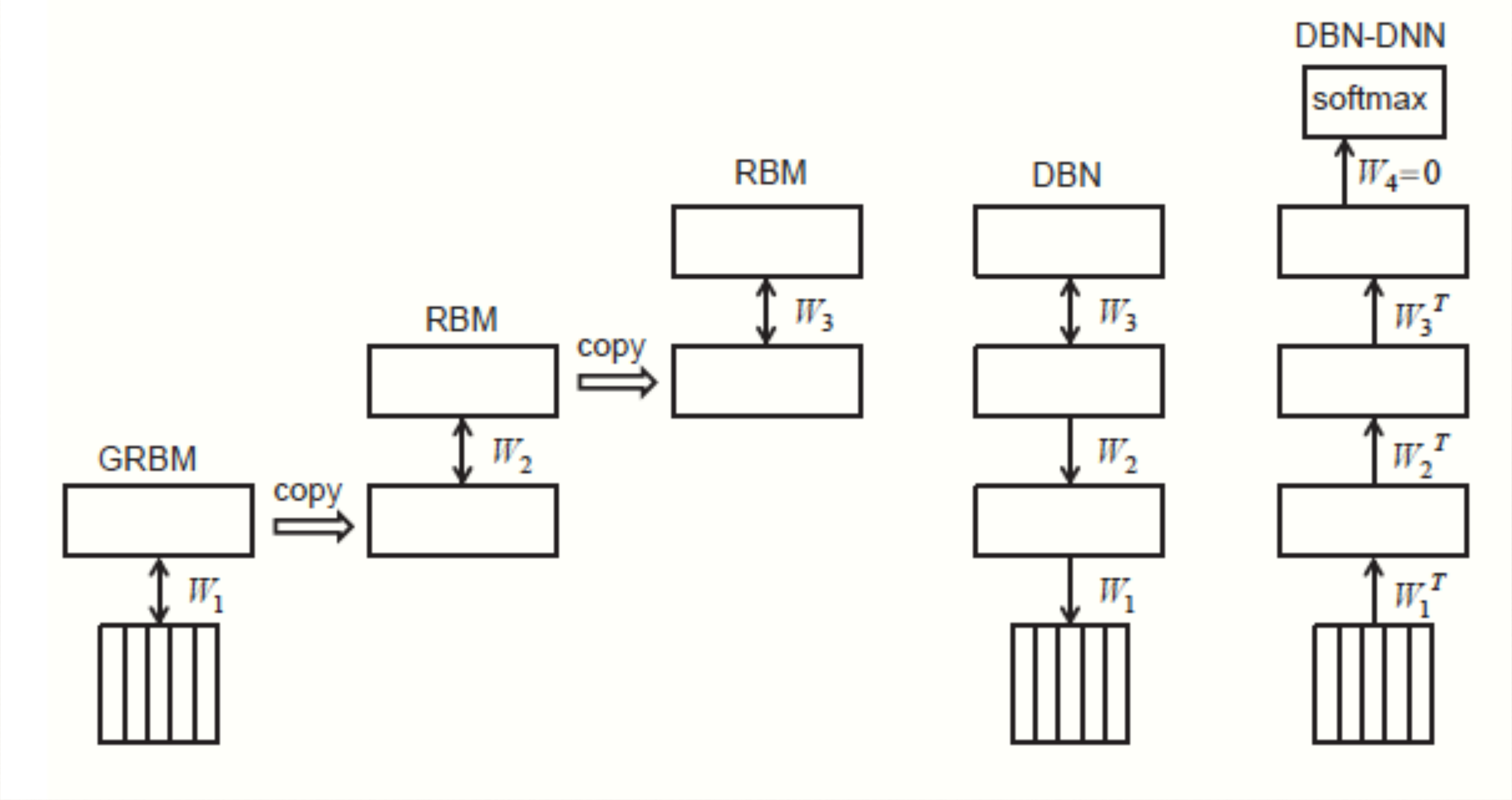
Earliest application was in dimensionality reduction [Hinton, 2006].

# Restricted Boltzmann Machines




**PCA** **RBM**

# DBN for DNN Initialization

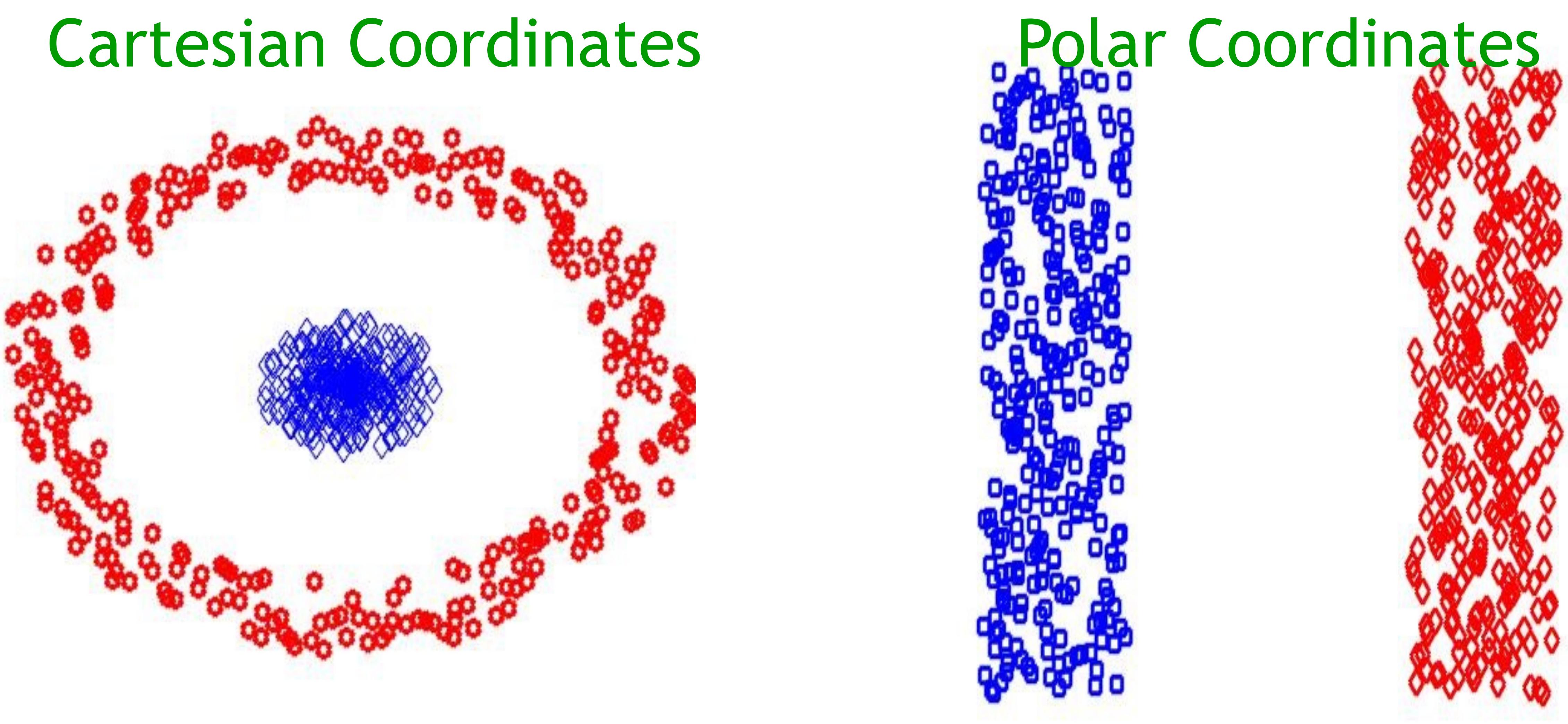



[Hinton, 2013]

# Understanding Deep Networks

# Representation Learning

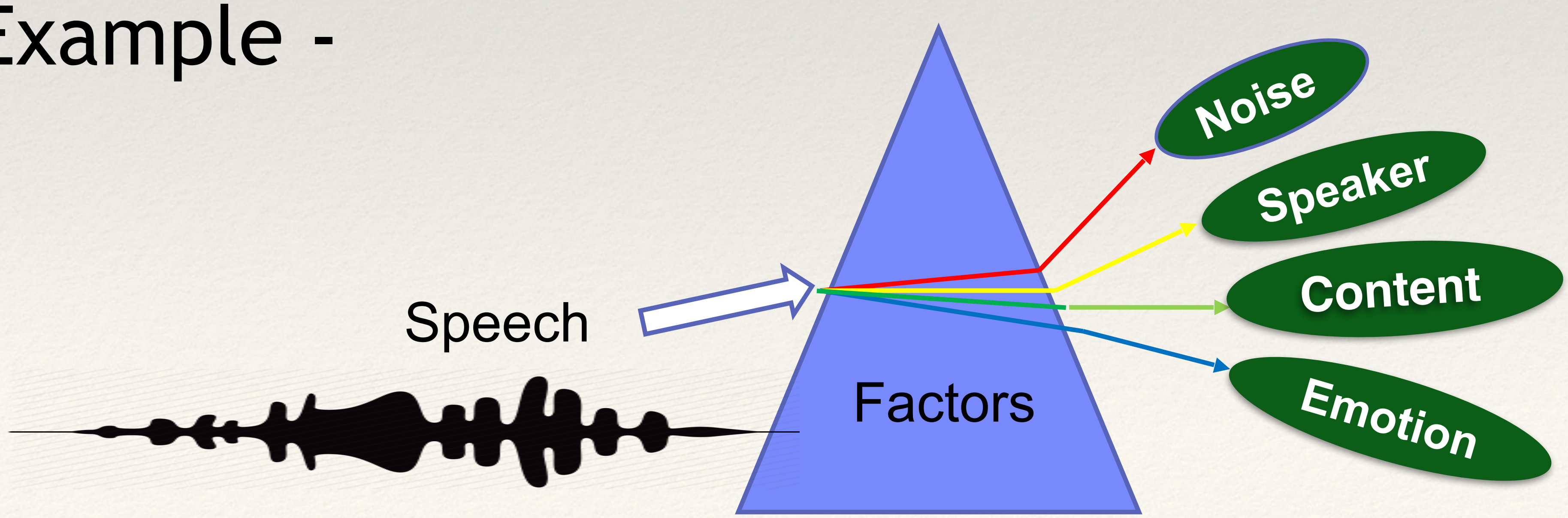
- The input data representation is one of most important components of any machine learning system.



# Representation Learning

- The input data representation is one of most important components of any machine learning system.
    - Extract features that enable classification while suppressing factors which are susceptible to noise.

- Finding the right representation for real world applications substantially challenging.
    - Example -



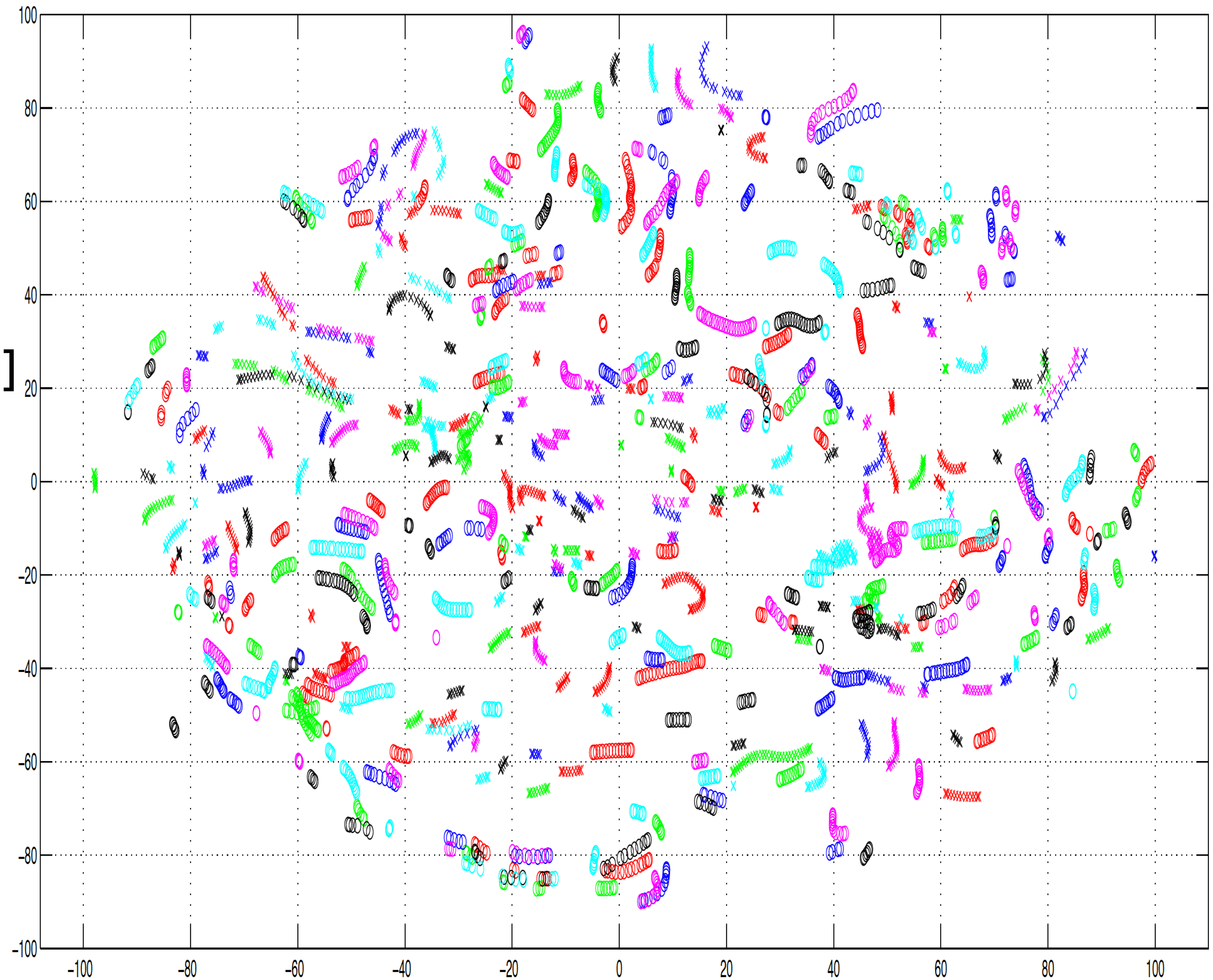
# Representation Learning

- The input data representation is one of most important components of any machine learning system.
    - Extract factors that enable classification while suppressing factors which are susceptible to noise.

- Finding the right representation for real world applications substantially challenging.
    - Deep learning solution - build complex representations from simpler representations.
    - The dependencies between these hierarchical representations are refined by the target.

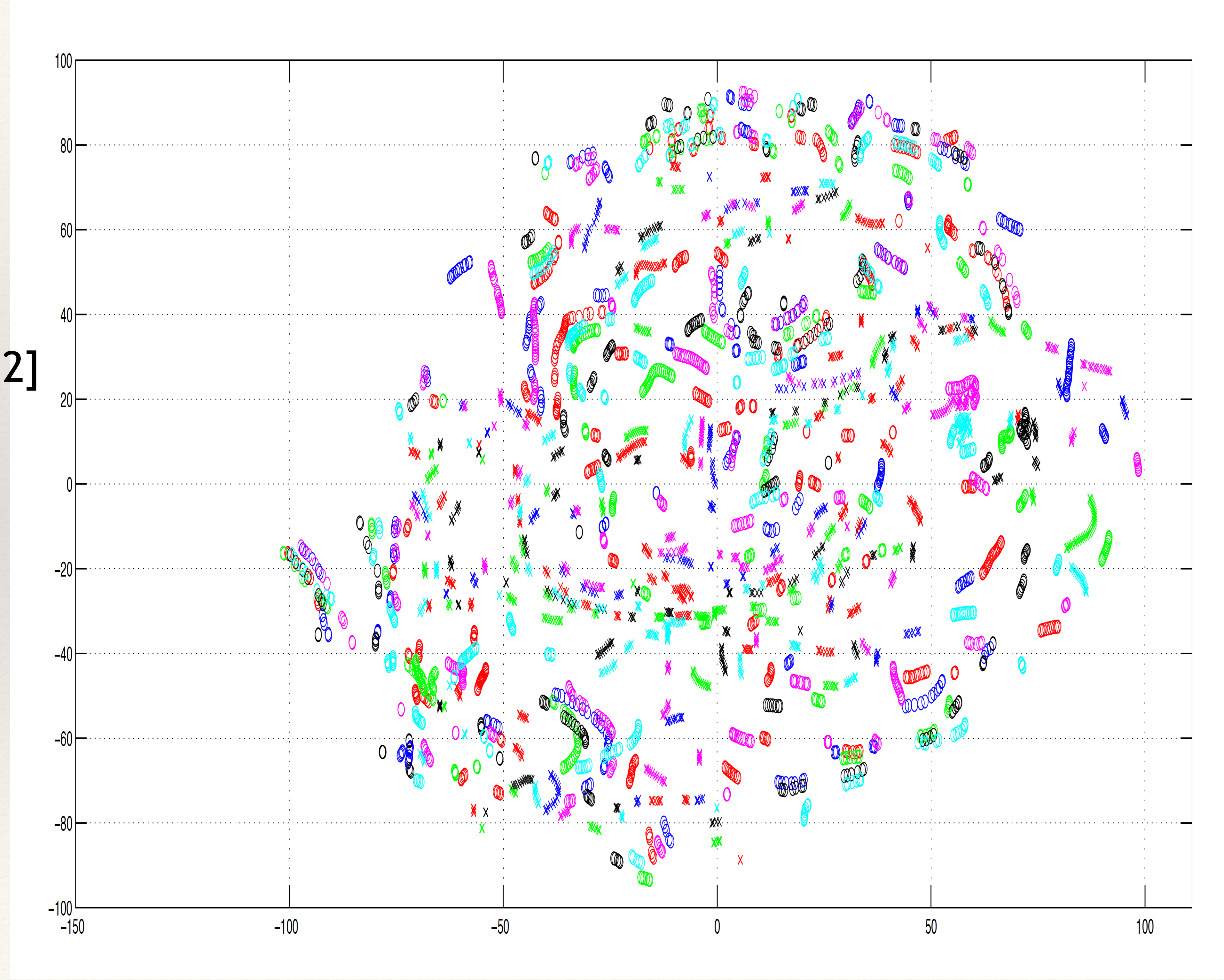
# Understanding DNNs for Speech

[Abdel Rahman, 2012]



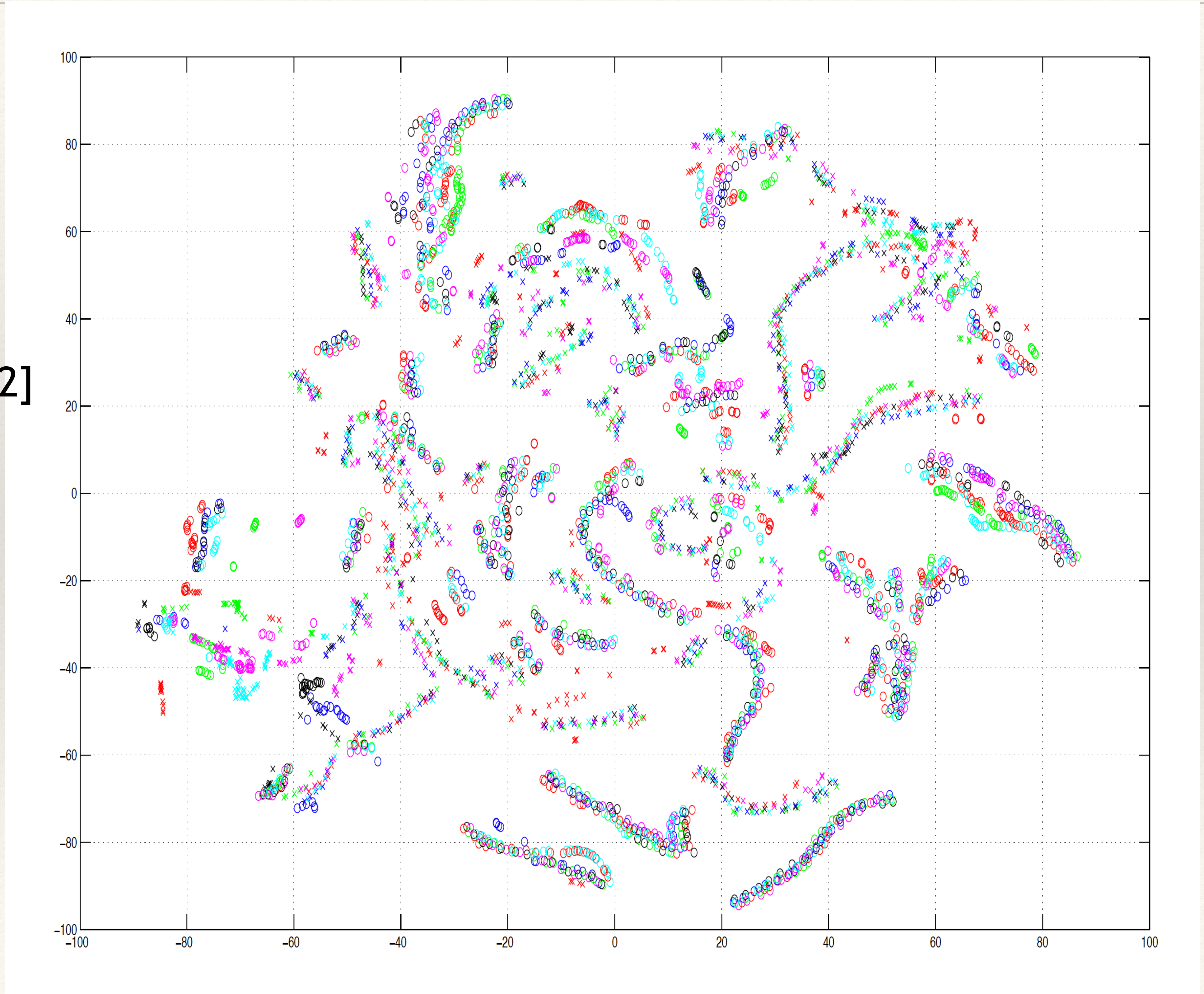
# Understanding DNNs for Speech

[Abdel Rahman, 2012]

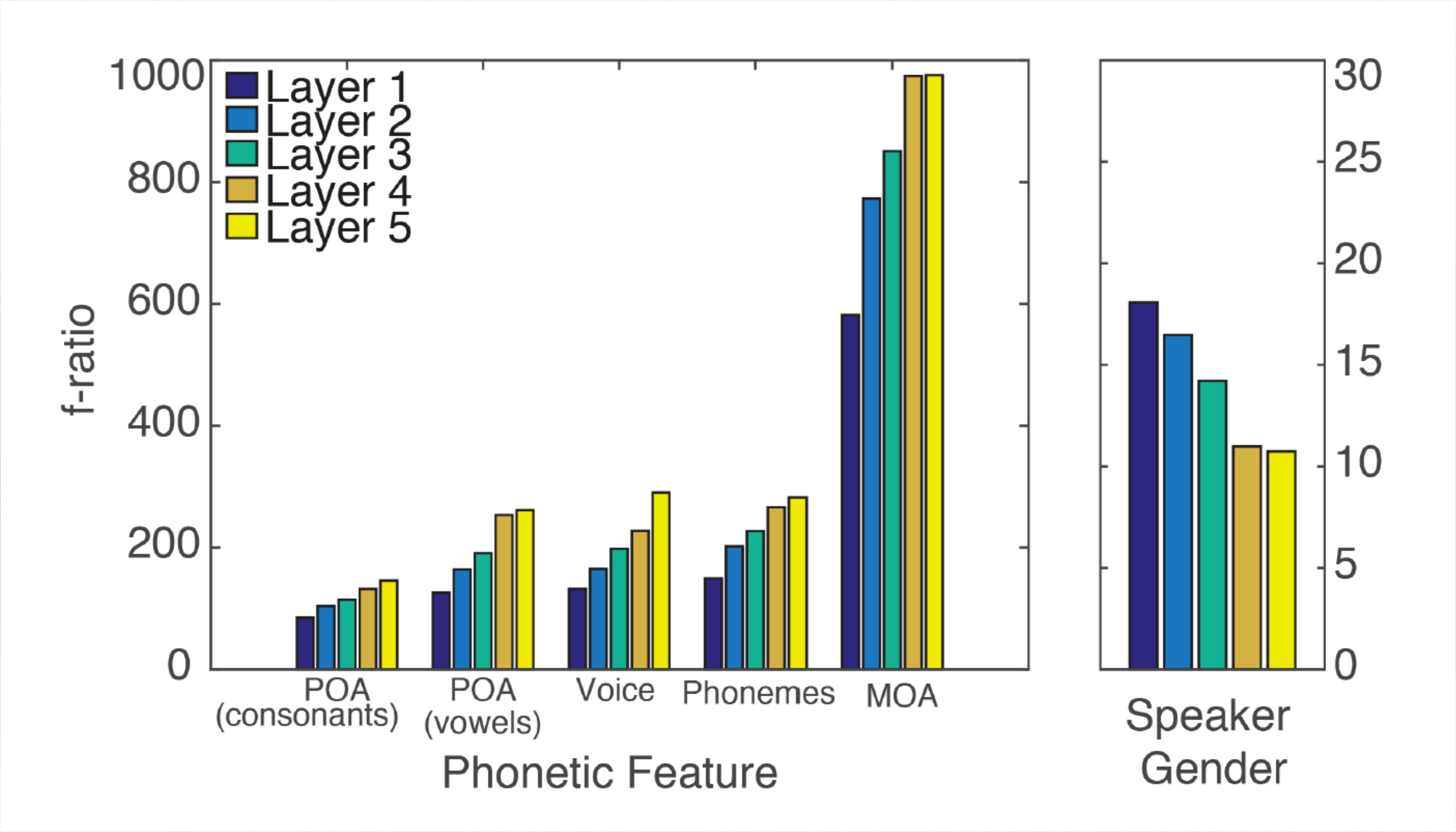


# Understanding DNNs for Speech

[Abdel Rahman, 2012]



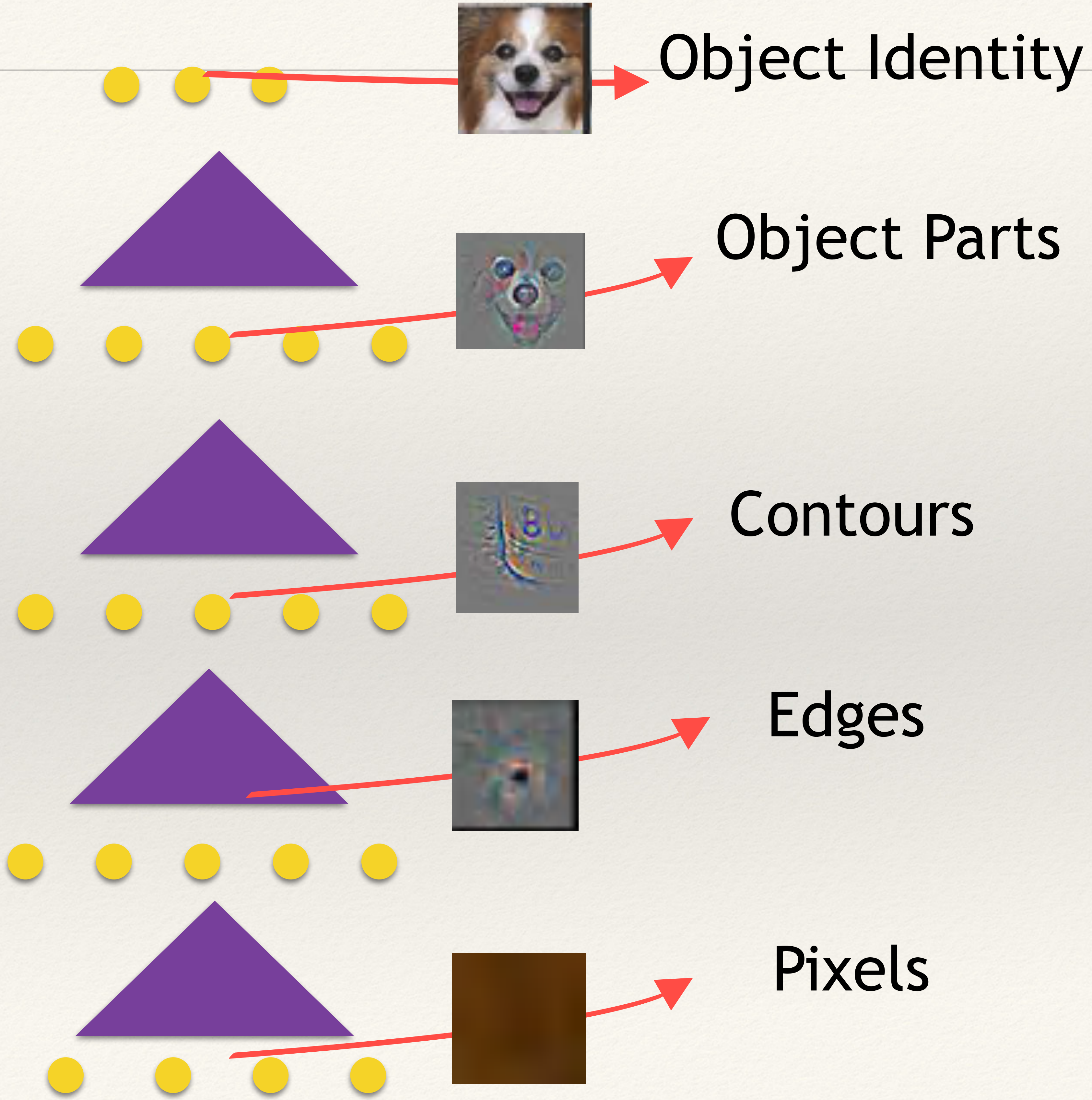
# Understanding DNNs for Speech




[Nagamine, 2015]

# Understanding DNNs for Image

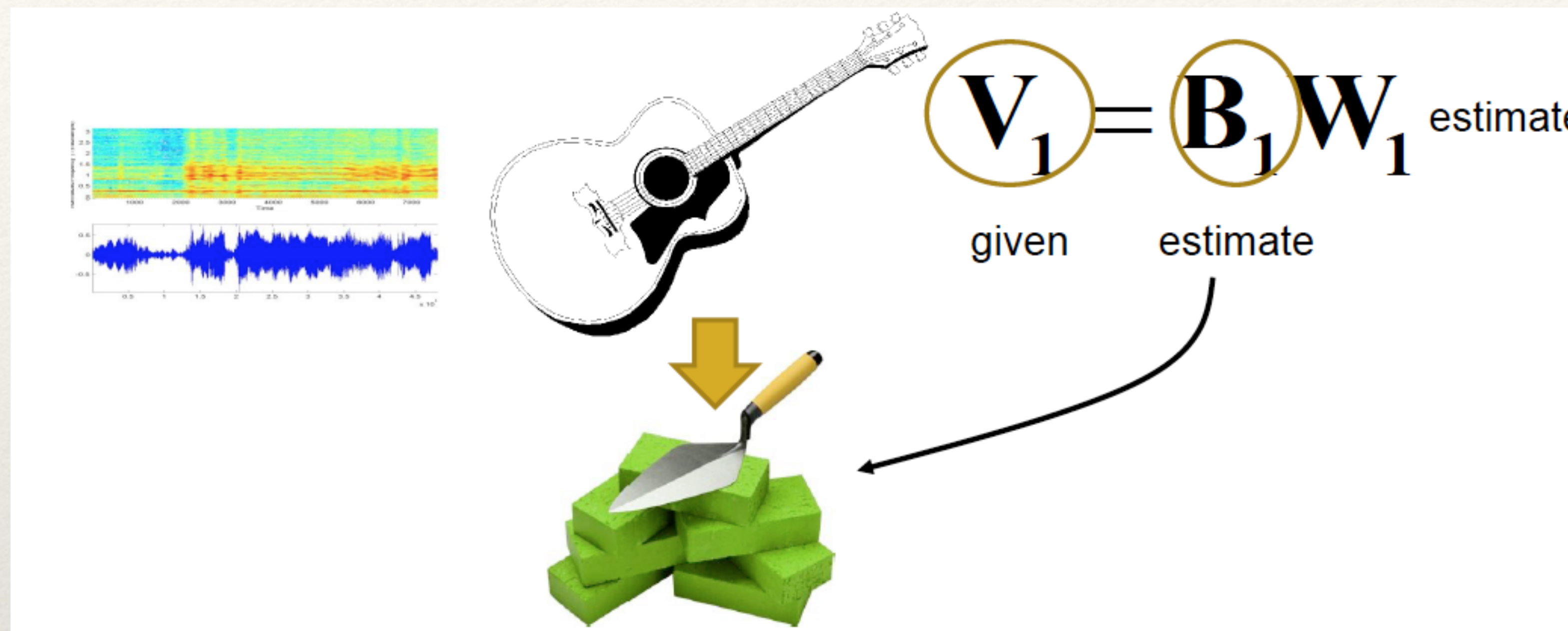
[Zeiler, 2014]



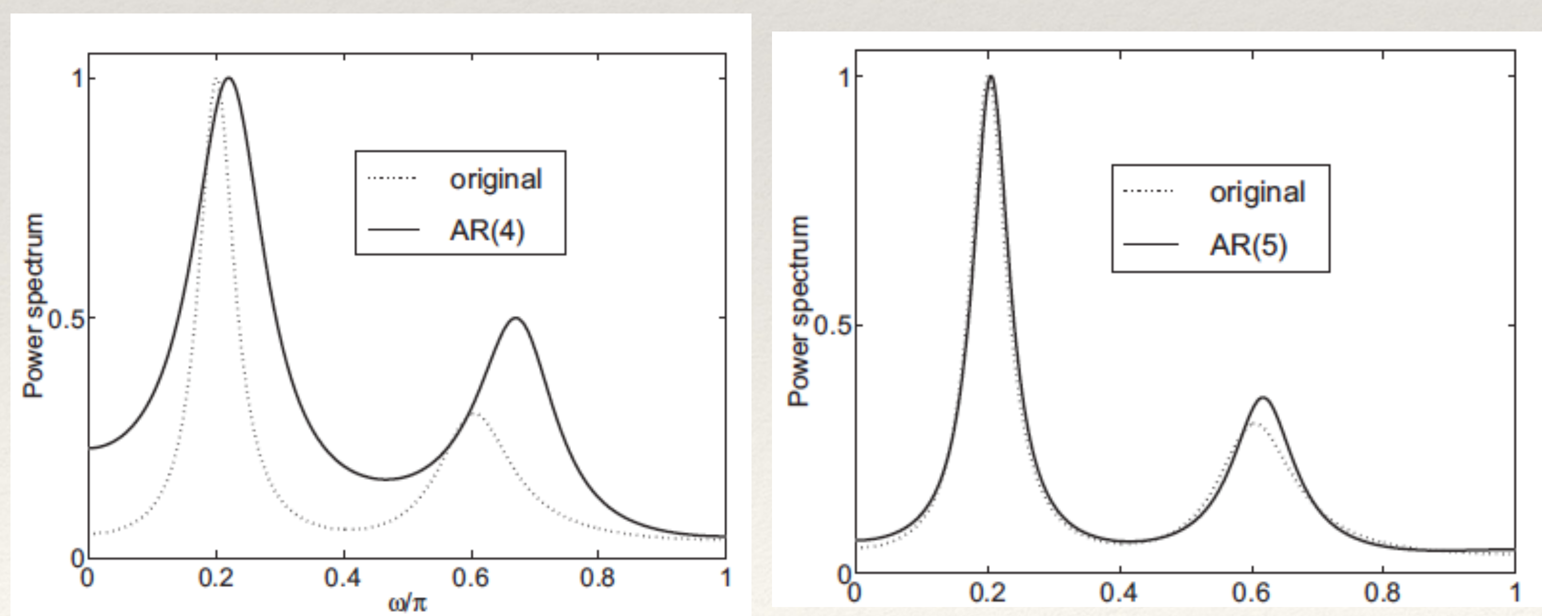
# Course Summary

# Introduction to Signal Processing

**NMF**

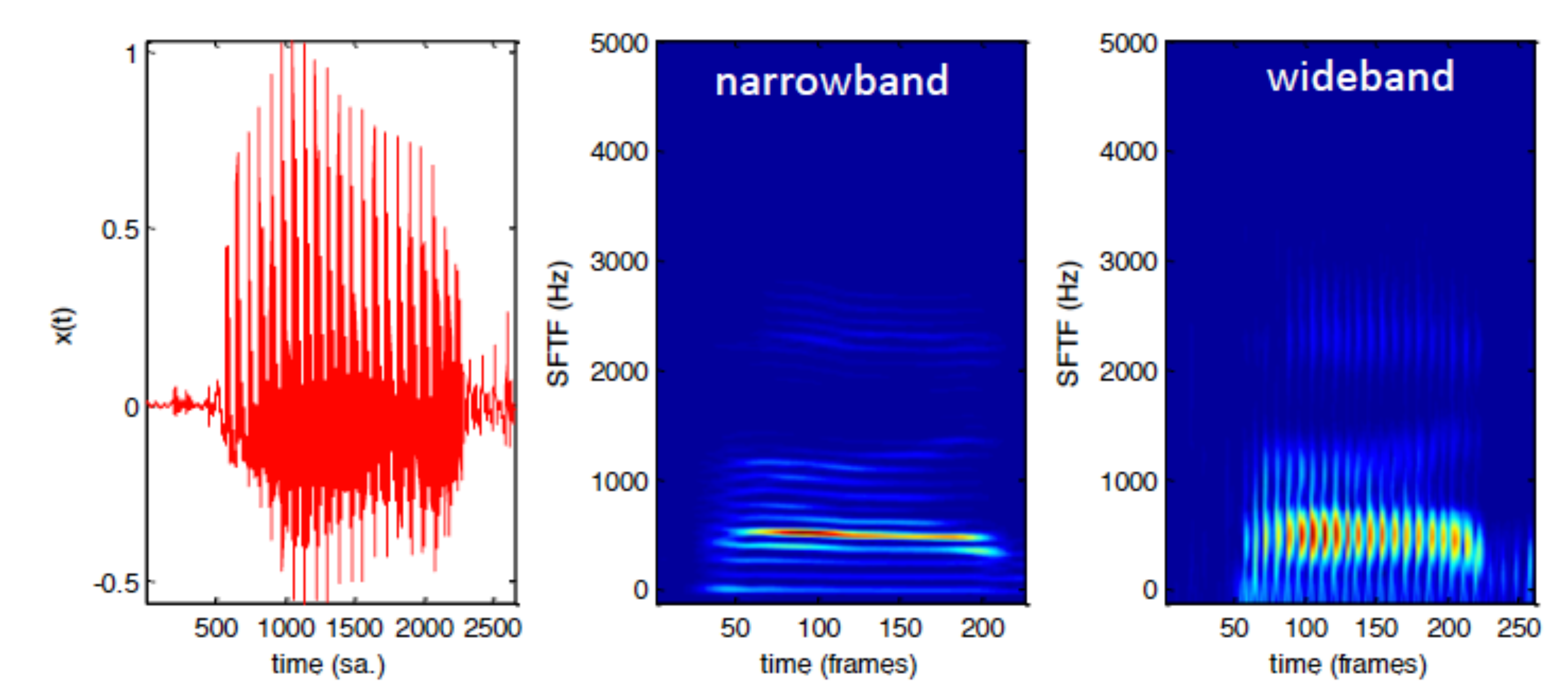



**Linear Prediction**

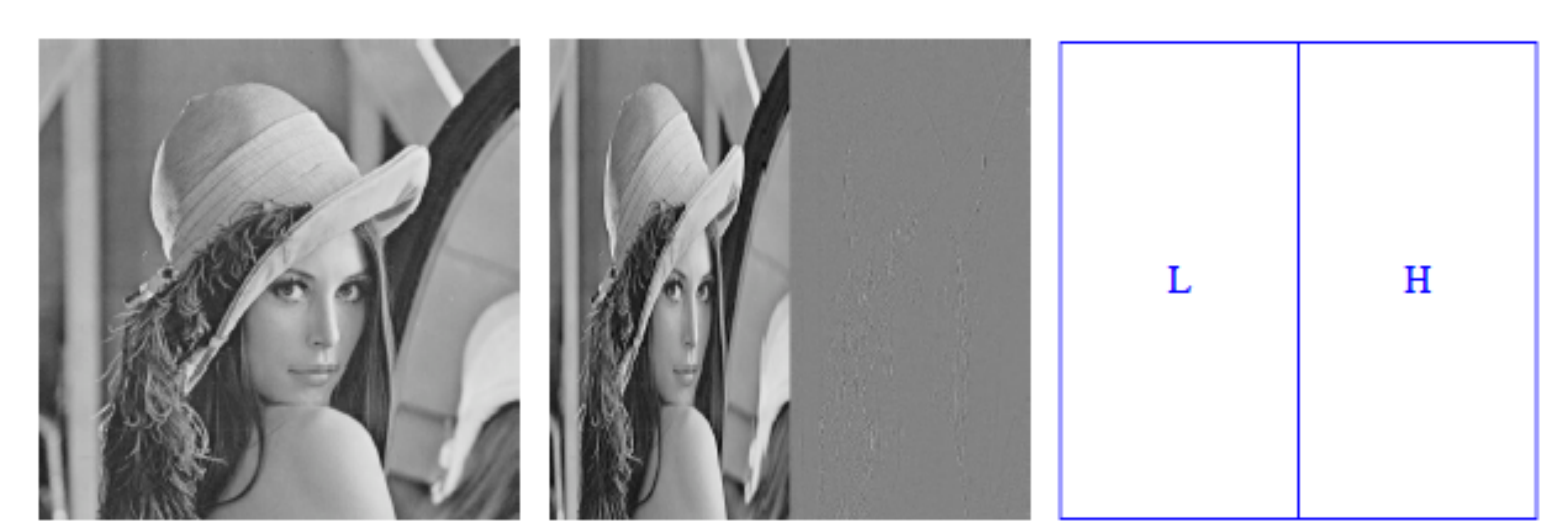


# Feature Extraction

Spectrogram

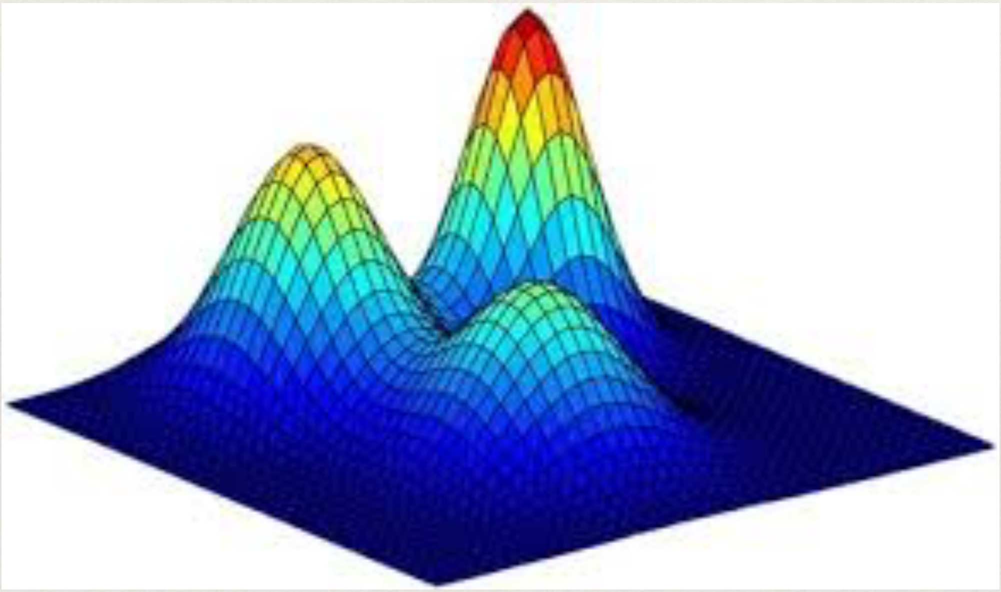



Wavelets

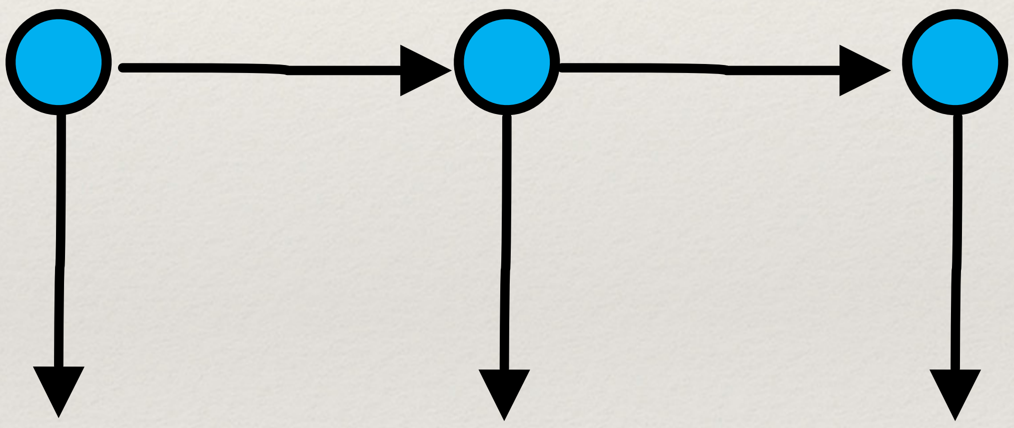


# Generative Models

**GMMs**



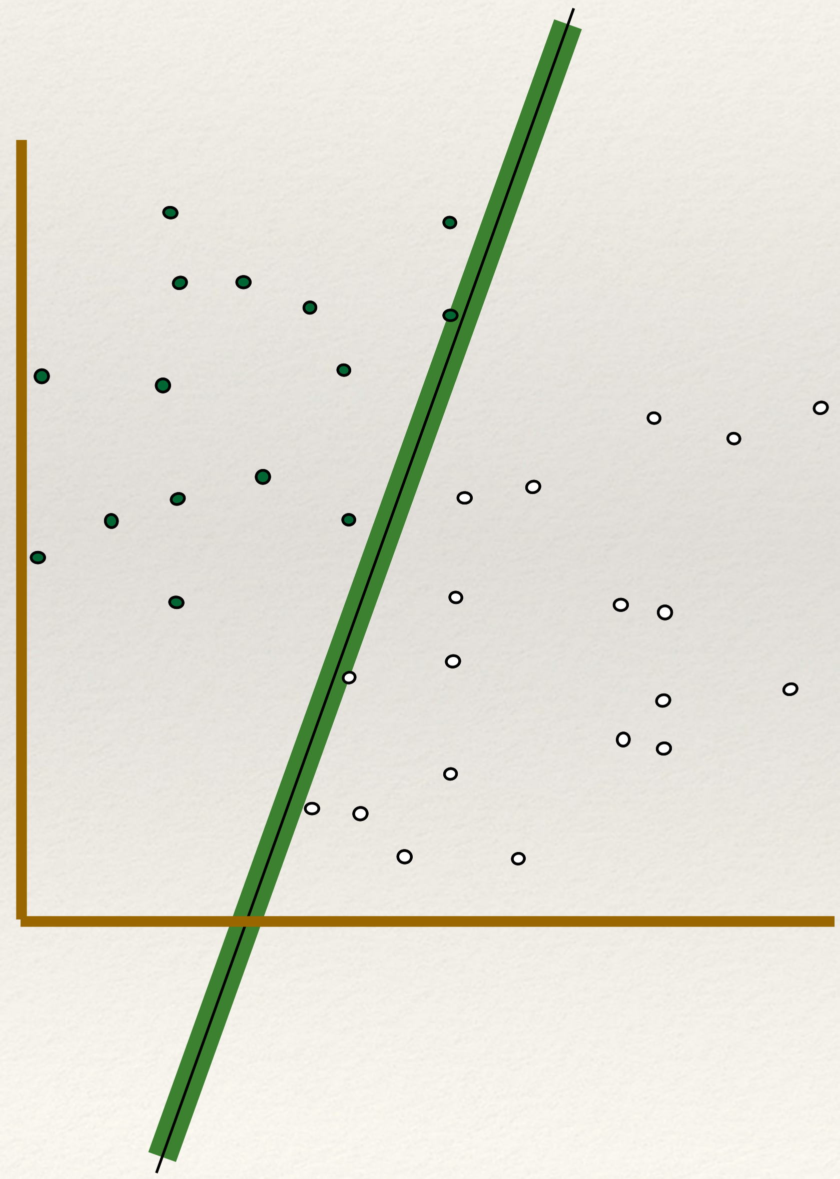
**HMMs**



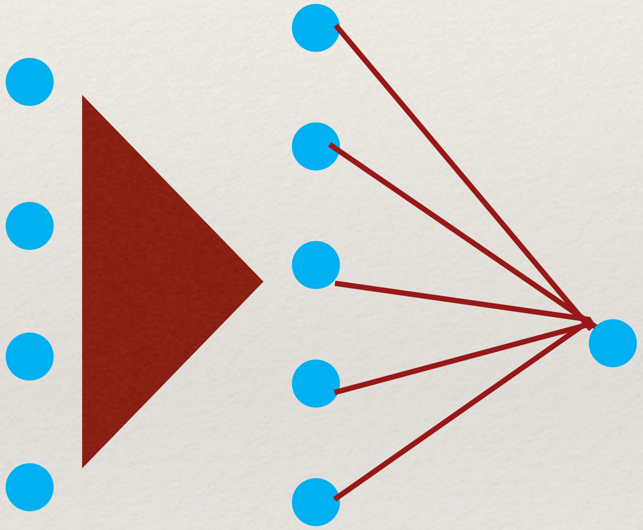
**Deep Generative Models**

# Discriminative Models

**SVMs**



**Neural Nets**



**Deep Networks**

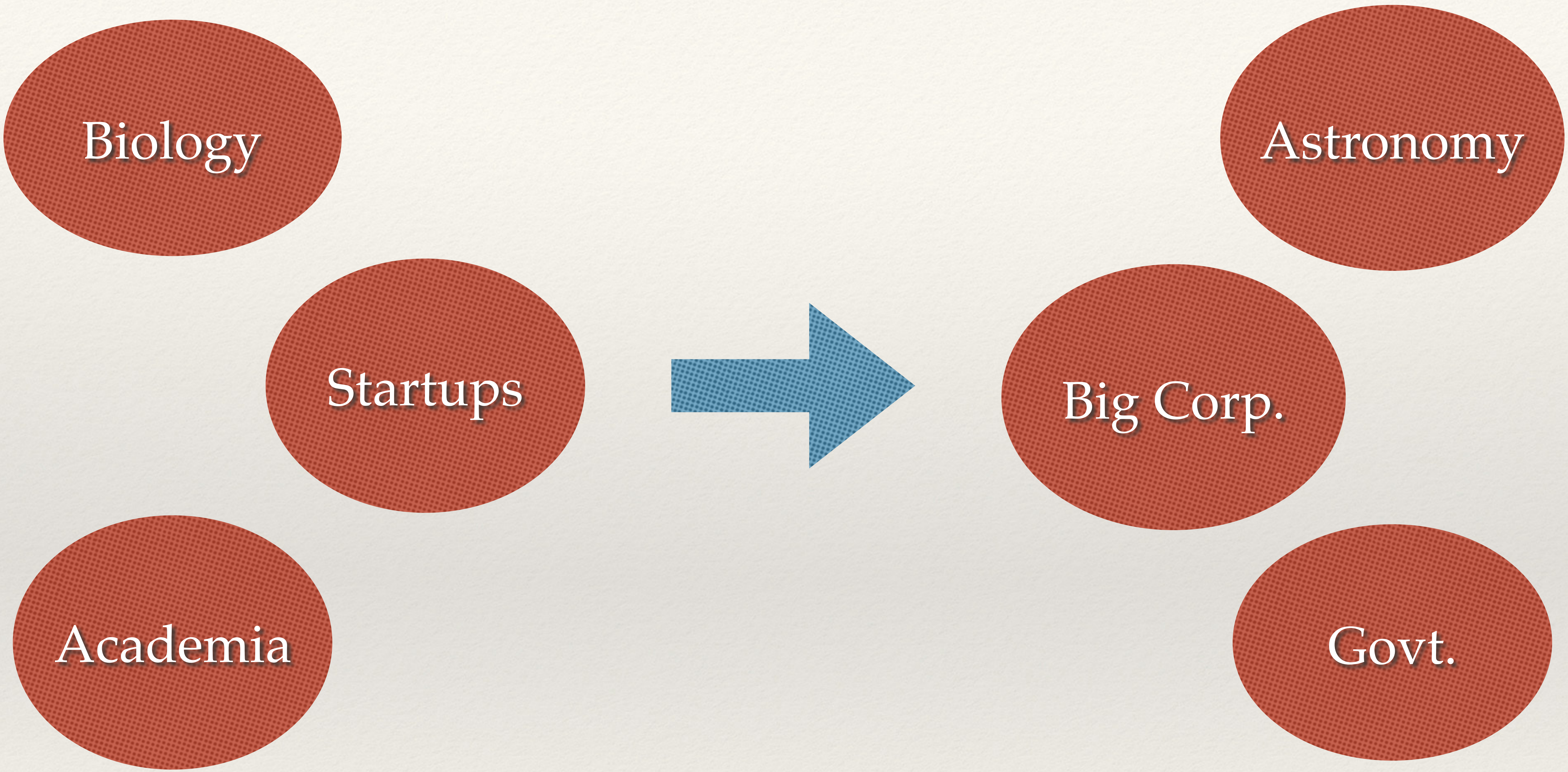
# Syllabus (From Webpage)

- Introduction to real world signals
- Feature extraction and front-end signal processing
- Basics of pattern recognition, Generative modeling - Gaussian and mixture Gaussia models, hidden Markov models, factor analysis and latent variable models.
- Discriminative modeling - support vector machines, neural networks and back propagation.
- Introduction to deep learning - convolutional and recurrent networks, pre-training and practical considerations in deep learning, understanding deep networks.
- Clustering methods and decision trees. Feature selection methods.
- Applications in computer vision and speech recognition.

# Syllabus (Based on Coverage)

- Introduction to real world signals
- Feature extraction and front-end signal processing
- Dimensionality Reduction Methods - PCA, LDA.
- Basics of pattern recognition, Generative modeling - Gaussian and mixture Gaussia models, hidden Markov models, factor analysis and latent variable models.
- Discriminative modeling - support vector machines, neural networks and back propagation.
- Introduction to deep learning - convolutional and recurrent networks, pre-training and practical considerations in deep learning, understanding deep networks.
- Deep Generative Models - RBMs and DBNs
- Applications in computer vision and speech recognition.

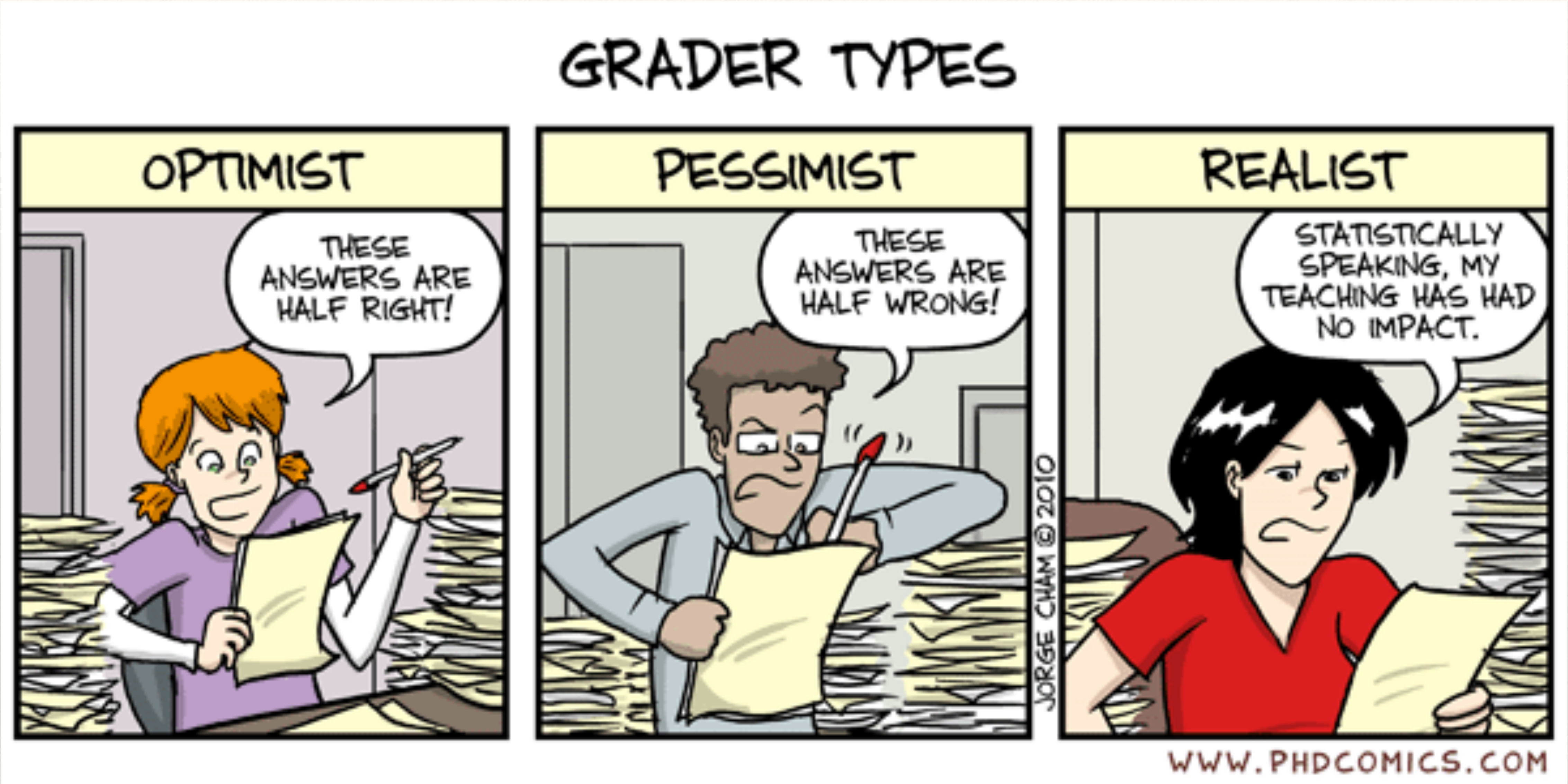
# Impact




**APPLE BUYS AI STARTUP THAT READS EMOTIONS IN FACES**

NVIDIA's Deep Learning Car Computer Selected by Volvo on Journey Toward a Crash-Free Future

# What have we achieved ?




But I learned a lot …