Cognitive models for emotion recognition: Big Data and Deep Learning

Big Data seminar

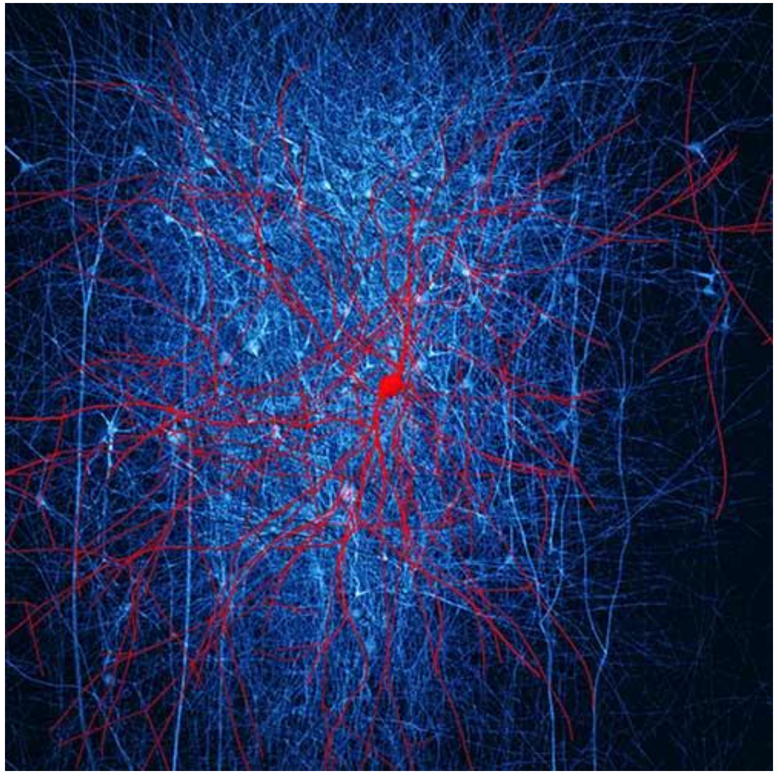
Presentation 10.14.15

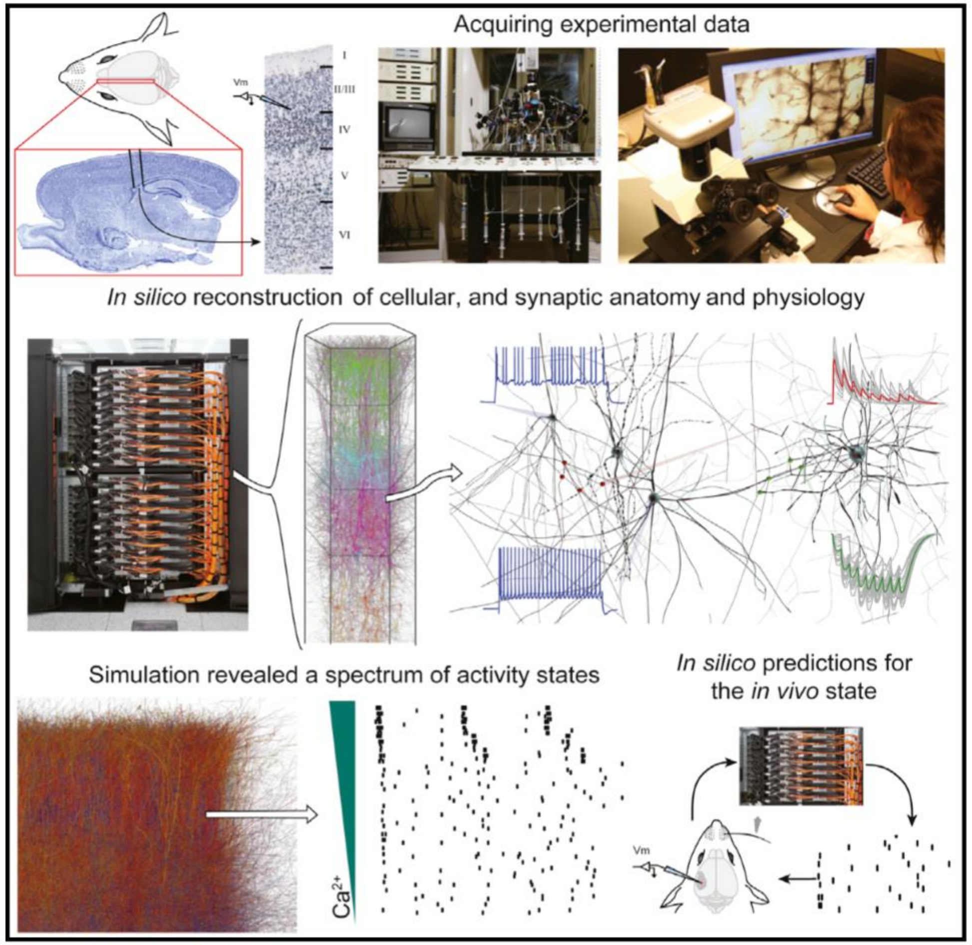
Outline

- Emotiv demo
- Data Acquisition
- Cognitive models for emotions recognition
- Big Data
- Deep Learning

Human Brain: the Big Data model
The most recent publication

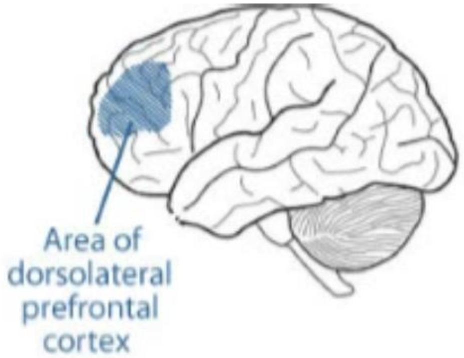
"Reconstruction and Simulation of Neocortical Microcircuitry", published in *Cell* on Oct. 8, 2015





It's a piece of rat brain containing about 30,000 neurons and 40 million synaptic connections.
There's nothing remarkable about it, except that it isn't real !
It's a digital reconstruction—a representation of a one-third cubic millimeter of rat neocortex—and it seems to work like the real thing.

Cerebral cortex



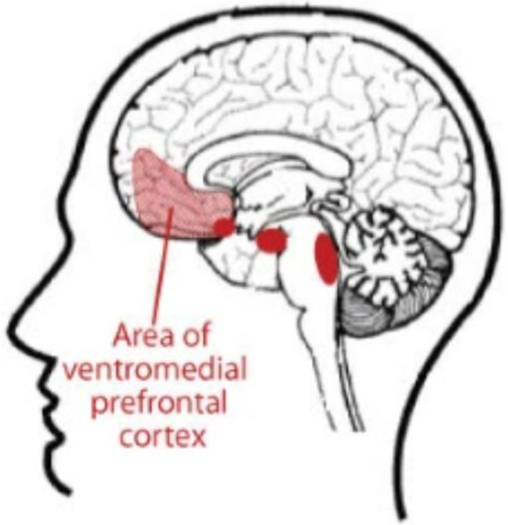
The newer portion of the cerebral **cortex** serves as the center of higher mental functions for humans.

Some Functions of the Ventromedial Prefrontal Cortex

Monitors brain and body information indicating emotional state

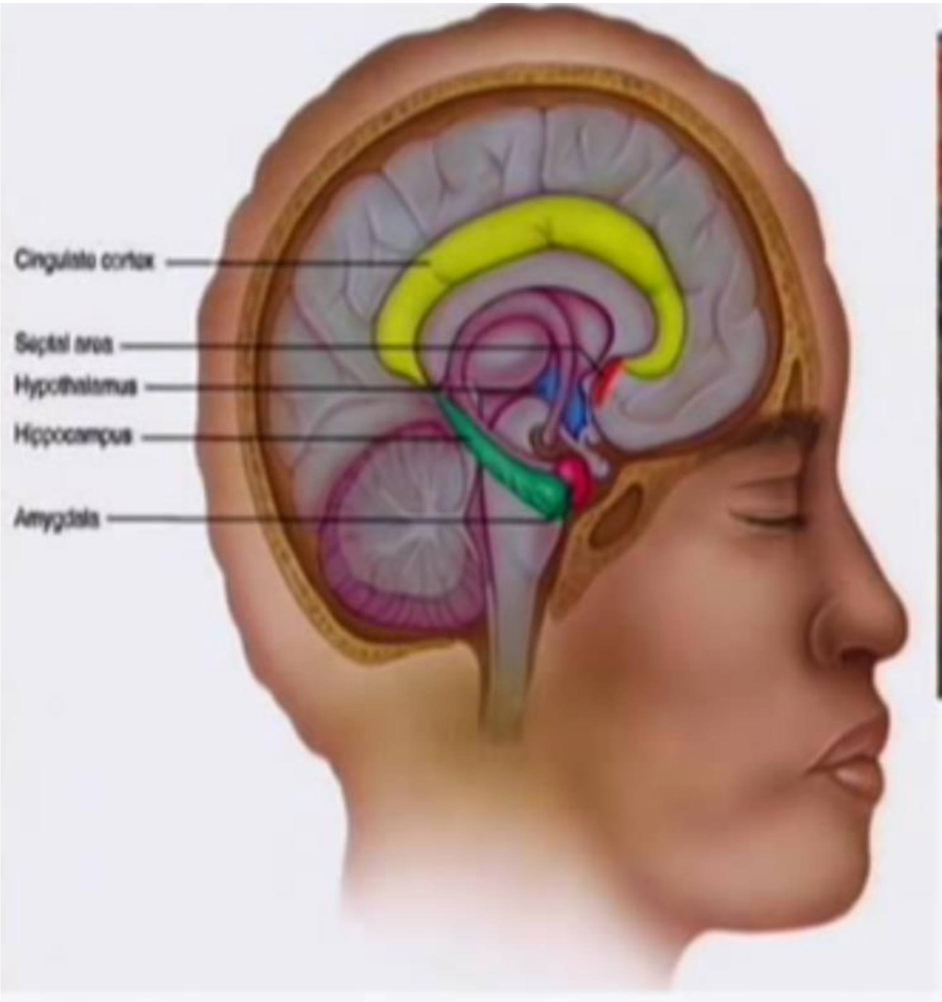
Helps to assess emotional relevance and emotional viability of prefrontal planning

Relays results of prefrontal processing to the amygdala, other emotion centers, and other brain areas.

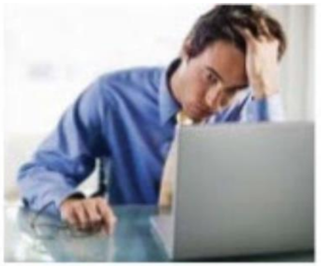


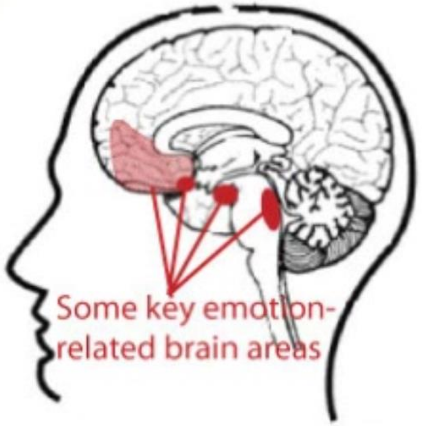
The neocortex contains some 100 billion cells.

Emotional brain



The spectrum of emotions

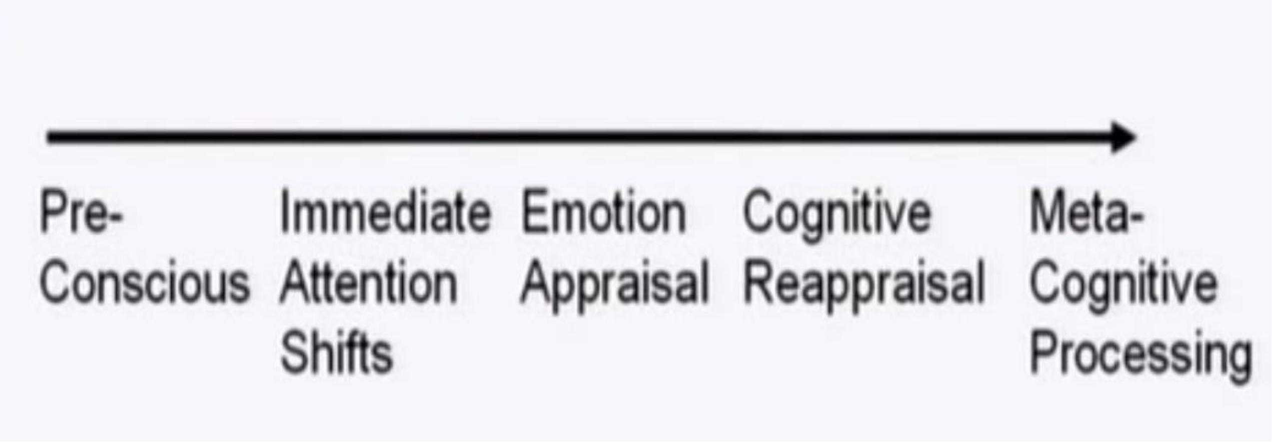




BACKGROUND FEELINGS

Irritation **Discontent** **Grumpiness**

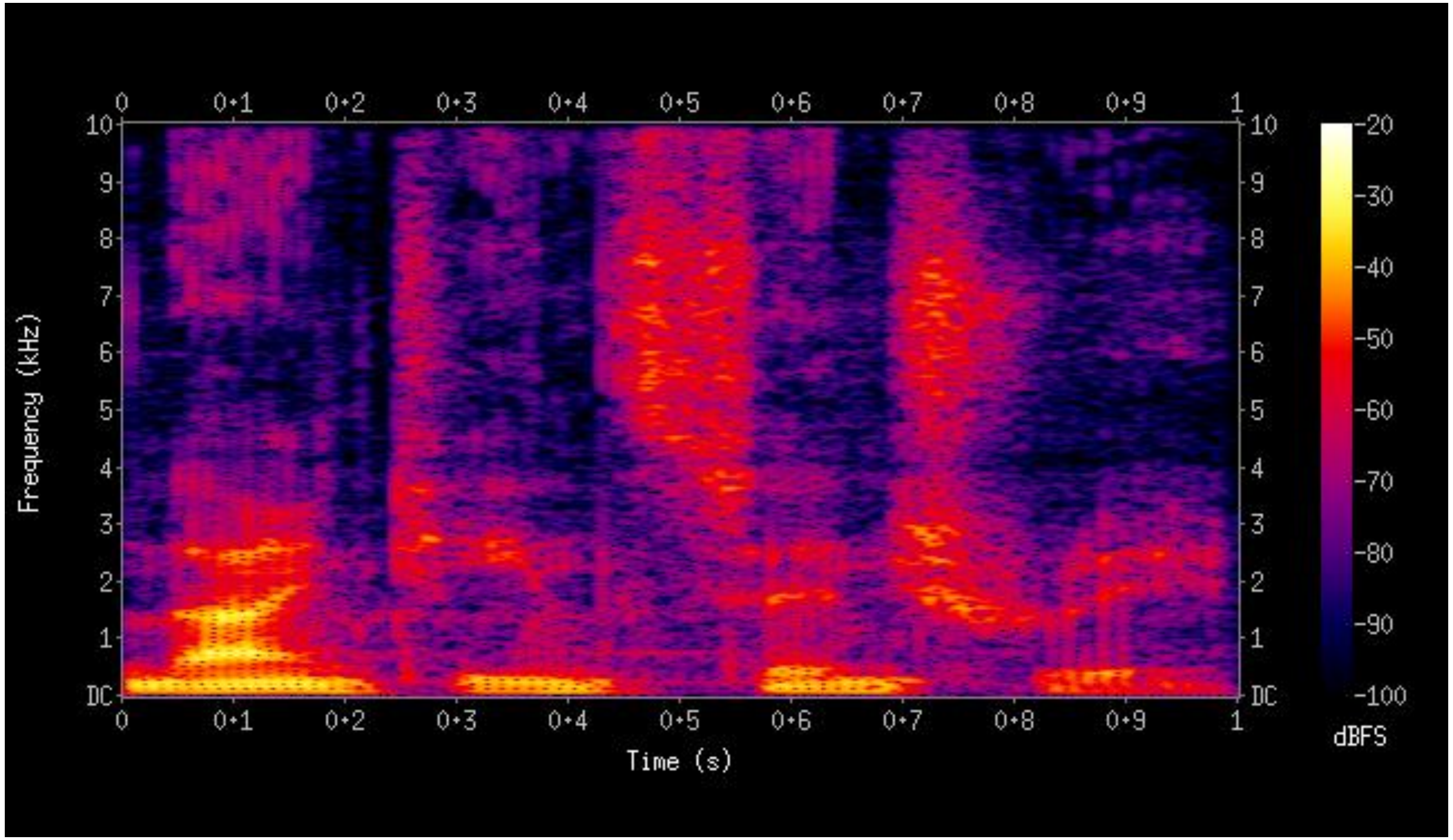
Stages of emotion regulation



MindDriver project: objectives

- Acquisition and processing of EEG data corresponding to different cognitive states and emotional responses of the human subjects.
- Used to feed the unsupervised and supervised learning models.
- Relies on the Deep Learning models with different levels of depth.
- Implemented using Big Data technologies.
- The task: Identify the most effective parameters of the learning procedure that would allow to reliably recognize such emotions, considered to be basic, as
    - anger, disgust, fear, happiness, sadness, and surprise,
- as well as their derivatives:
    - amusement, contempt, contentment, embarrassment, excitement, guilt, pride in achievement, relief, satisfaction, sensory pleasure, and shame.

Signal processing

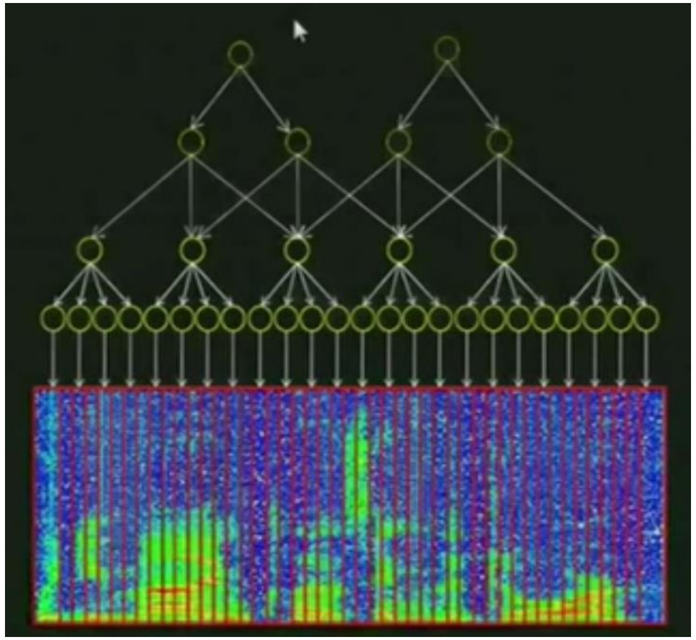


- FFT and filter banks
- Walsh spectrum
- Fractals
- Wavelets

Resulted in spectrograms with a number of temporal windows

The research objective: identify the method most adequate for emotion recognition and mapping onto the brain cortex

Deep Learning as a DBN



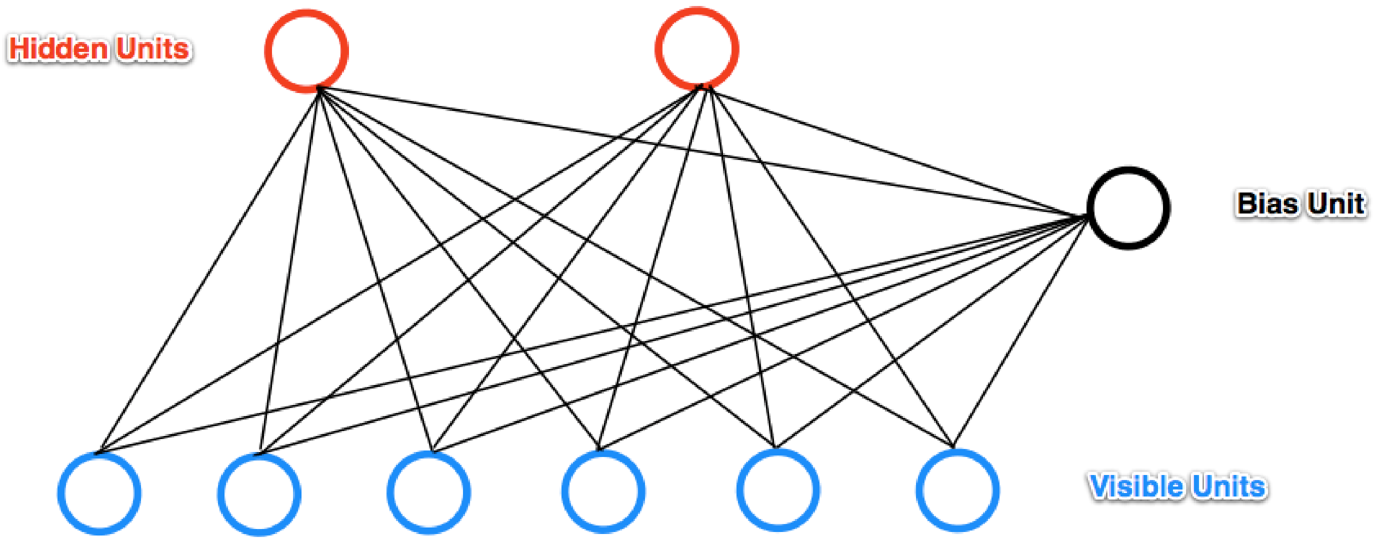
- A fragment of a Deep Belief Network
- Unsupervised learning
- Spectrogram as an input (*visual nodes*)
- Stacked layers
- Supervised layer as a final one
- Parallel solutions at each stage (Map-Reduce)
- Backpropagation at the last stage

Deep Learning classification

- Deep networks for unsupervised or generative learning, which are intended to capture high-order correlation of the observed or visible data for pattern analysis or synthesis purposes when no information about target class labels is available.
- Deep networks for supervised learning, which are intended to directly provide discriminative power for pattern classification purposes, often by characterizing the posterior distributions of classes conditioned on the visible data. Target label data are always available in direct or indirect forms for such supervised learning. They are also called discriminative deep networks.
- Hybrid deep networks, where the goal is discrimination which is assisted, often in a significant way, with the outcomes of generative or unsupervised deep networks.

Restricted Boltzmann Machine

- The most widely used hybrid deep architecture
- A special type of Markov random field that has one layer of (typically Bernoulli) stochastic hidden units and one layer of (typically Bernoulli or Gaussian) stochastic visible or observable units



RBM: theory and method

The joint distribution *p*(**v**,**h***;ϑ*) over the visible units *v* and hidden units *h*, given the model parameters *ϑ*, is defined in terms of an energy function *E*(*v,h;ϑ*) of

$$p(\mathbf{v},\mathbf{h};\theta) = \frac{\exp(-E(\mathbf{v},\mathbf{h};\theta))}{Z}, \qquad Z = \sum_{\mathbf{v}}\sum_{\mathbf{h}}\exp(-E(\mathbf{v},\mathbf{h};\theta))$$

The marginal probability that the model assigns to a visible vector v is

$$p(\mathbf{v};\theta) = \frac{\sum_{\mathbf{h}}\exp(-E(\mathbf{v},\mathbf{h};\theta))}{Z}$$

The energy function is defined as

$$E(\mathbf{v},\mathbf{h};\theta) = -\sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}v_i h_j - \sum_{i=1}^{I} b_i v_i - \sum_{j=1}^{J} a_j h_j.$$

The conditional probabilities can be efficiently calculated as

$$p(h_j = 1|\mathbf{v};\theta) = \sigma\left(\sum_{i=1}^{I} w_{ij}v_i + a_j\right),$$

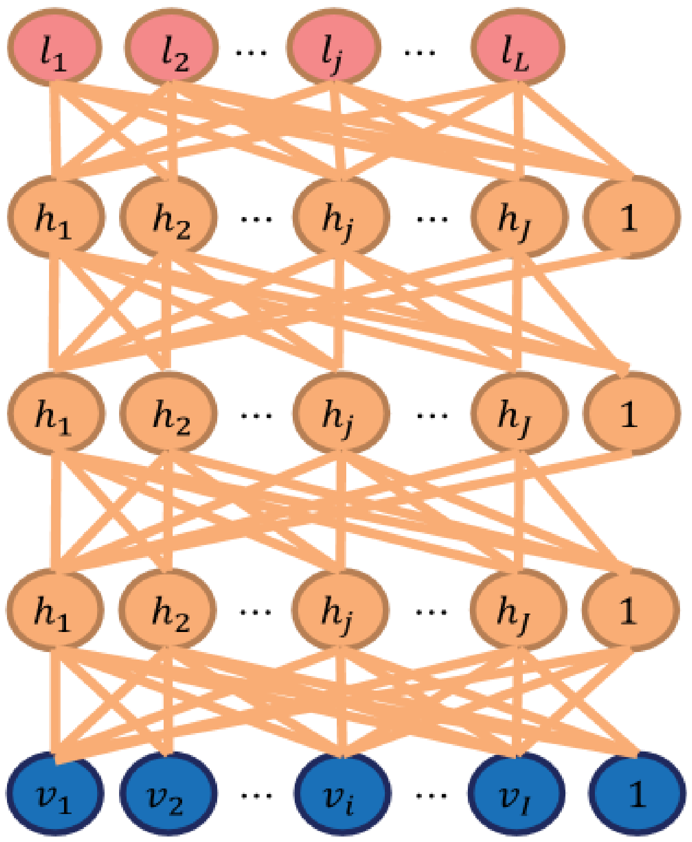
$$p(v_i = 1|\mathbf{h};\theta) = \sigma\left(\sum_{j=1}^{J} w_{ij}h_j + b_i\right),$$

$$\sigma(x) = 1/(1+\exp(-x)).$$

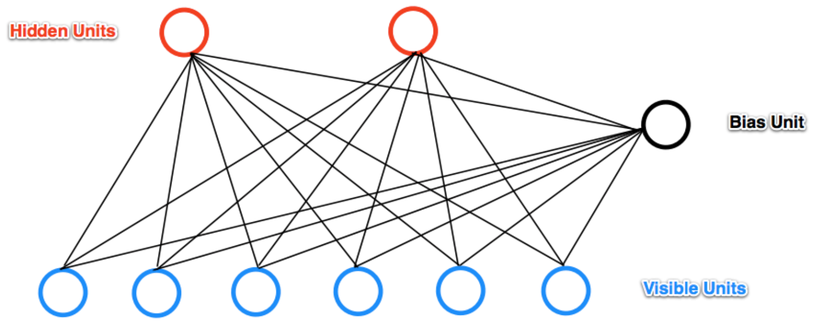
Deep Learning: Stacking the RBM’s

- Stacking a number of the RBMs learned layer by layer from bottom up gives rise to a DBN.
- The stacking procedure is as follows. After learning a RBM with binary features such as spectral bins indexed for different electrodes, we treat the activation probabilities of its hidden units as the data for training the Bernoulli Bernoulli RBM one layer up.
- The activation probabilities of the second layer Bernoulli-Bernoulli RBM are then used as the visible data input for the third-layer Bernoulli-Bernoulli RBM, and so on.
- It has been shown that the stacking procedure improves a variational lower bound on the likelihood of the training data under the composite model.
- That is, this greedy procedure achieves approximate maximum likelihood learning.
- This learning procedure is unsupervised and requires no class label.

The Deep Belief Network – Deep Neural Network (DBN-DNN) architecture



Training example (fictiouos)



Six visible nodes (V1..V6): theta rhythm presence (1) and otherwise (0) at different electrode locations

Two hidden nodes to differentiate unknown emotional states

Training samples: A(1,1,1,0,0,0), B(1,0,1,0,0,0), C(1,1,1,0,0,0), D(0,1,1,1,0,0), E(0,0,1,1,1,0), F(0,0,1,1,1,0)

The network learned the following weights:

	Bias Unit	Hidden 1	Hidden 2
Bias Unit	-0.08257658	-0.19041546	1.57007782
V1	-0.82602559	-7.08986885	4.96606654
V2	-1.84023877	-5.18354129	2.27197472
V3	3.92321075	2.51720193	4.11061383
V4	0.10316995	6.74833901	-4.00505343
V5	-0.97646029	3.25474524	-5.59606865
V6	-4.44685751	-2.81563804	-2.91540988

In the testing mode, a sample (0,0,0,1,1,) is tested. It turns Hidden 1 on and Hidden 2 off. Interpretation?