VS 265 - Neural Computation

Bruno A. Olshausen, Instructor
Office: 570 Evans
baolshausen@berkeley.edu

Brian Cheung, Mayur Mudigonda, GSI’s
Office: 567 Evans
bcheung, mudigonda @berkeley.edu
Class meets TTH 3:30-5  
Room 560, Evans Hall

Weekly Matlab assignments (60% of grade)

Final Project (40% of grade)

Readings:  
  Handouts  
  Hertz, Krogh & Palmer, *Introduction to the Theory of Neural Computation*  
  Dayan & Abbott, *Theoretical Neuroscience*  
  MacKay, *Information Theory, Inference and Learning Algorithms*

Wiki page:  
  http://redwood.berkeley.edu/wiki/VS265

Class email list:  vs265-students@lists.berkeley.edu
Schedule (for next few weeks):

Week 1 (Aug. 28): Introduction

Week 2 (Sept. 2, 4): Neuron models, Perceptron model

Week 3 (Sept. 9, 11): guest lectures

Week 4 (Sept. 16, 18): Multilayer perceptrons

Week 5 (Sept. 23, 25): Unsupervised learning and PCA

Week 6 (Sept. 30, Oct. 2): Competitive learning

Week 7 (Oct. 7, 9): Plasticity and cortical maps
Readings for this week
(available on the wiki)

Today:


For Tuesday:

- Linear neuron models (handout)
- Linear time-invariant systems and convolution (handout)
- Simulating differential equations (handout)
What have brain scans and single-unit recording taught us about the *computations* underlying perception and cognition?
After 50 years of concerted research efforts...

- machines are still incapable of solving simple perceptual or motor control tasks.

- there is little understanding of how neurons interact to process sensory information or to produce actions.

We are missing something fundamental on both fronts: we are ignorant of the underlying principles governing perception and action.
What’s so hard about it?
Among the most challenging scientific questions of our time are the corresponding analytic and synthetic problems: How does the brain function? Can we design a machine which will simulate a brain?

-- *Automata Studies*, 1956
Subgoal for July

Analysis of scenes consisting of non-overlapping objects from the following set:

balls

bricks with faces of the same or different colors or textures

cylinders.

Each face will be of uniform and distinct color and/or texture.

Background will be homogeneous.

Extensions for August

The first priority will be to handle objects of the same sort but with complex surfaces and backgrounds, e.g. cigarette pack with writing and bands of different color, or a cylindrical battery.

Then extend class of objects to objects like tools, cups, etc.
The Lighthill debate (1973)

http://www.aiai.ed.ac.uk/events/lighthill1973/

Sir James Lighthill
Our first foray into Artificial Intelligence was a program that did a credible job of solving problems in college calculus. Armed with that success, we tackled high school algebra; we found, to our surprise, that it was much harder. Attempts at grade school arithmetic, involving the concept of numbers, etc., provide problems of current research interest. An exploration of the child’s world of blocks proved insurmountable, except under the most rigidly constrained circumstances. It finally dawned on us that the overwhelming majority of what we call intelligence is developed by the end of the first year of life.

--Minksy, 1977
Even ‘simple’ nervous systems can exhibit profound visual intelligence.
...problem solving behavior, language, expert knowledge and application, and reason, are all pretty simple once the essence of being and reacting are available. That essence is the ability to move around in a dynamic environment, sensing the surroundings to a degree sufficient to achieve the necessary maintenance of life and reproduction. This part of intelligence is where evolution has concentrated its time—it is much harder.

The theory reported here clearly demonstrates the feasibility and fruitfulness of a quantitative statistical approach to the organization of cognitive systems. By the study of systems such as the perceptron, it is hoped that those fundamental laws of organization which are common to all information handling systems, machines and men included, may eventually be understood.” -- Frank Rosenblatt

### The approach of David Marr

<table>
<thead>
<tr>
<th>Computational theory</th>
<th>Representation and algorithm</th>
<th>Hardware implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?</td>
<td>How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?</td>
<td>How can the representation and algorithm be realized physically?</td>
</tr>
</tbody>
</table>

*Figure 1–4.* The three levels at which any machine carrying out an information-processing task must be understood.
The approach of David Marr

**Viewer centred**
- **Input Image**
  - Perceived intensities
- **Primal Sketch**
  - Zero crossings, blobs, edges, bars, ends, virtual lines, groups, curves boundaries.
- **2 1/2-D Sketch**
  - Local surface orientation and discontinuities in depth and in surface orientation

**Object centred**
- **3-D Model Representation**
  - 3-D models hierarchically organised in terms of surface and volumetric primitives
Natural images are full of ambiguity
Natural images are full of ambiguity
What do these patterns depict?

(from Kersten & Yuille, 2003)
Vision as inference
Nervous systems are difficult to penetrate
1 mm$^2$ of cortex contains 100,000 neurons
Anatomy of a synapse
Are there principles?

“God is a hacker”
– Francis Crick

“...their (neurons’) apparently erratic behavior was caused by our ignorance, not the neuron’s incompetence.”
– H.B. Barlow (1972)
Principles of optics govern the design of eyes
What are the principles that govern the operation of this system?
squares of perpendicular distance, and least absolute deviations, the third of which is more robust against outliers. The standard deviations for the slope and intercept were estimated directly for the first method and by bootstrap for the last two methods. Bootstrap may help detect outliers in the data because, when they are left out from a same-size resample, the correlation coefficient often increases, which could be exploited to improve estimation. Systematic bias caused by outliers was not detected in Fig. 2.

3. Theory of Scaling

Our analysis rests on two assumptions. First, we assume that each small piece of cortex of unit area, regardless of its thickness and the overall brain size, sends and receives about the same total cross-sectional area of long-distance connection fibers to and from other cortical regions. Second, we assume that the global geometry of the cortex minimizes the average length of the long-distance fibers.

The second assumption follows from Ramon y Cajal's principle for conservation of space, conduction time, and cellular materials (Chap. V in ref. 15). This principle has been explored more recently as the principle of minimal axon length (16–18). Consistent with previous observations on the basic uniformity of the cortex (19–21), the first assumption is supported loosely by the evidence that the total number of neurons beneath a unit cortical surface area is about $10^5/20862$ mm$^2$ across different cortical regions for several species, from mouse to human (22) (after shrinkage correction). But there are exceptions, including the higher density in striate cortex of primates (22, 23), the lower density in dolphin cortex (24), and the variability observed in cat cortex (25). The number of axons leaving or entering the gray–white boundary per unit cortical area should be compared.

Fig. 2. Cortical white and gray matter volumes of various species are related by a power law that spans five to six orders of magnitude. Most data points are based on measurement of a single adult animal. The line is the least squares fit, with a slope around $1.23/0.01$ (mean SD). The average and median deviations of the white matter volumes from the regression line are, respectively, 18% and 13% on a linear scale. Sources of data: If the same species appeared in more than one source below, the one mentioned earlier was used. All 38 species in table 2 in ref. 3 were taken, including 23 primates, 2 tree shrews, and 13 insectivores. Another 11 species were taken from table 2 in ref. 8, including 3 primates, 2 carnivores, 4 ungulates, and 2 rodents. Five additional species came from table 1 in ref. 11, including 1 elephant and 4 cetaceans. The data point for the mouse ($G = 112$ mm$^3$ and $W = 13$ mm$^3$) was based on ref. 30, and that for the rat ($G = 425$ mm$^3$ and $W = 59$ mm$^3$) was measured from the serial sections in a stereotaxic atlas (42). The estimates for the fisherman bat ($Noctilio leporinus$, $G = 329$ mm$^3$ and $W = 43$ mm$^3$) and the flying fox ($Pteropus lylei$, $G = 2,083$ mm$^3$ and $W = 341$ mm$^3$) were based on refs. 43 and 44, with the ratios of white and gray matters estimated roughly from the section photographs in the papers. The sea lion data ($Zalophus californianus$, $G = 113,200$ mm$^3$ and $W = 56,100$ mm$^3$) were measured from the serial sections at the website given in the legend to Fig. 1, with shrinkage correction.

$$
\log_{10} W = (1.23 \pm 0.01) \log_{10} G - (1.47 \pm 0.04)
$$

$$
r = 0.998
$$
Recurrent computation is pervasive throughout cortex
Computational principles

• Efficient coding
• Unsupervised learning
• Bayesian inference
• Dynamical systems
• Prediction
Experiment $\rightarrow$ Theory $\rightarrow$ Experiment
“Von Neumann bottleneck”
McClelland, Rumelhart & Hinton (ca. 1985):

“...a number of different pieces of information must be kept in mind at once.”

“To articulate these intuitions, we and others have turned to a class of models we call Parallel Distributed Processing (PDP) models. These models assume that information processing takes place through the (simultaneous) interactions of a large number of simple processing elements called units, each sending excitatory and inhibitory signals to the other units.”