Machine Learning Department
School of Computer Science
Carnegie Mellon University

Vector Spaces

Matt Gormley
Lecture 4
September 10, 2018

Q&A

Q: Why did the readings on the course Schedule page not match the lectures?

A: Sorry! That was a copy/paste error. Now resolved.

Q: Is there an official audit option for this course?

A: There was not before today. If you are interested, please talk to me after class. (Today is the university deadline for switching to audit.)

Q&A

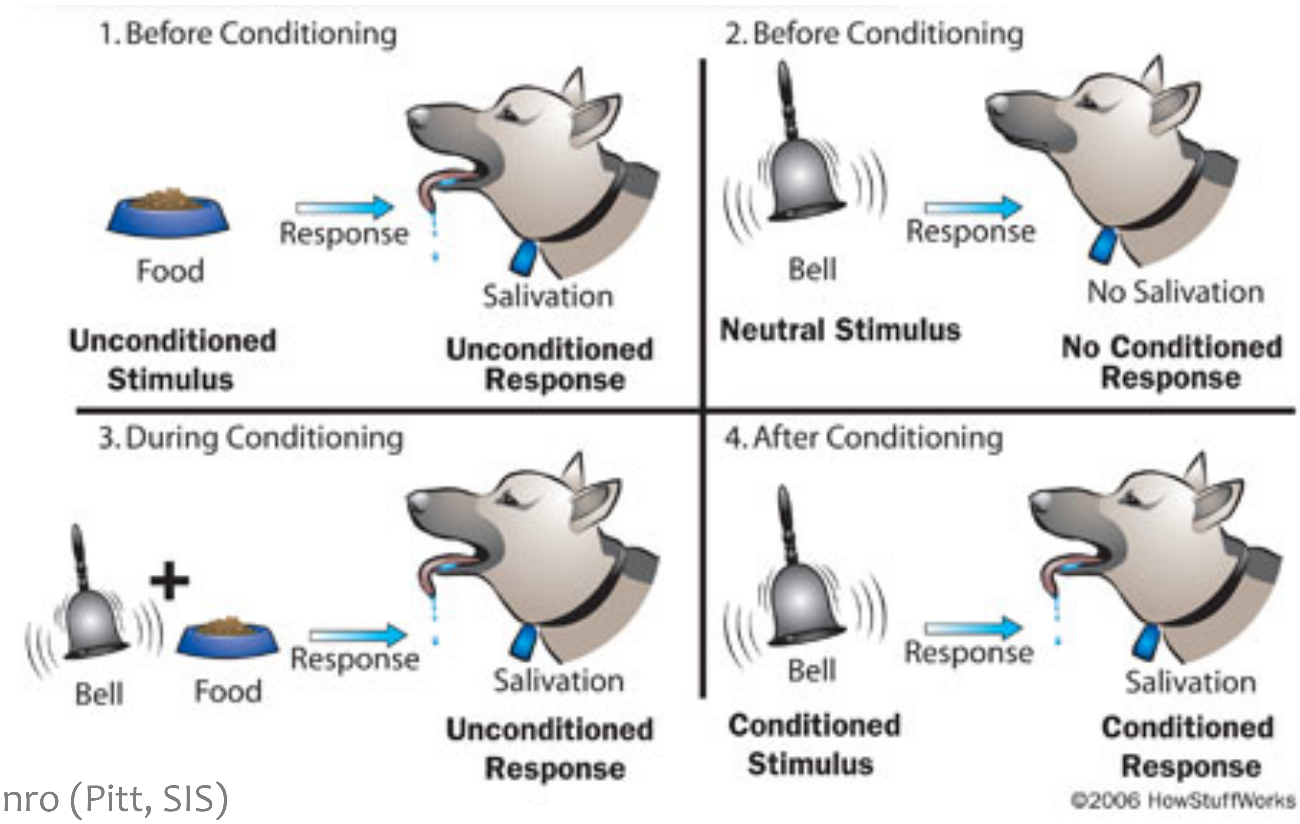
MOTIVATION: LINEAR ALGEBRA

Learning

Slide from Paul Munro (Pitt, SIS)

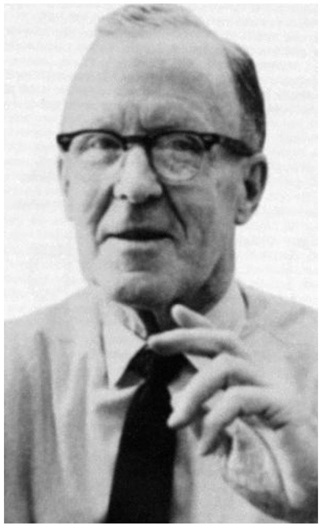
Neural substrate of learning

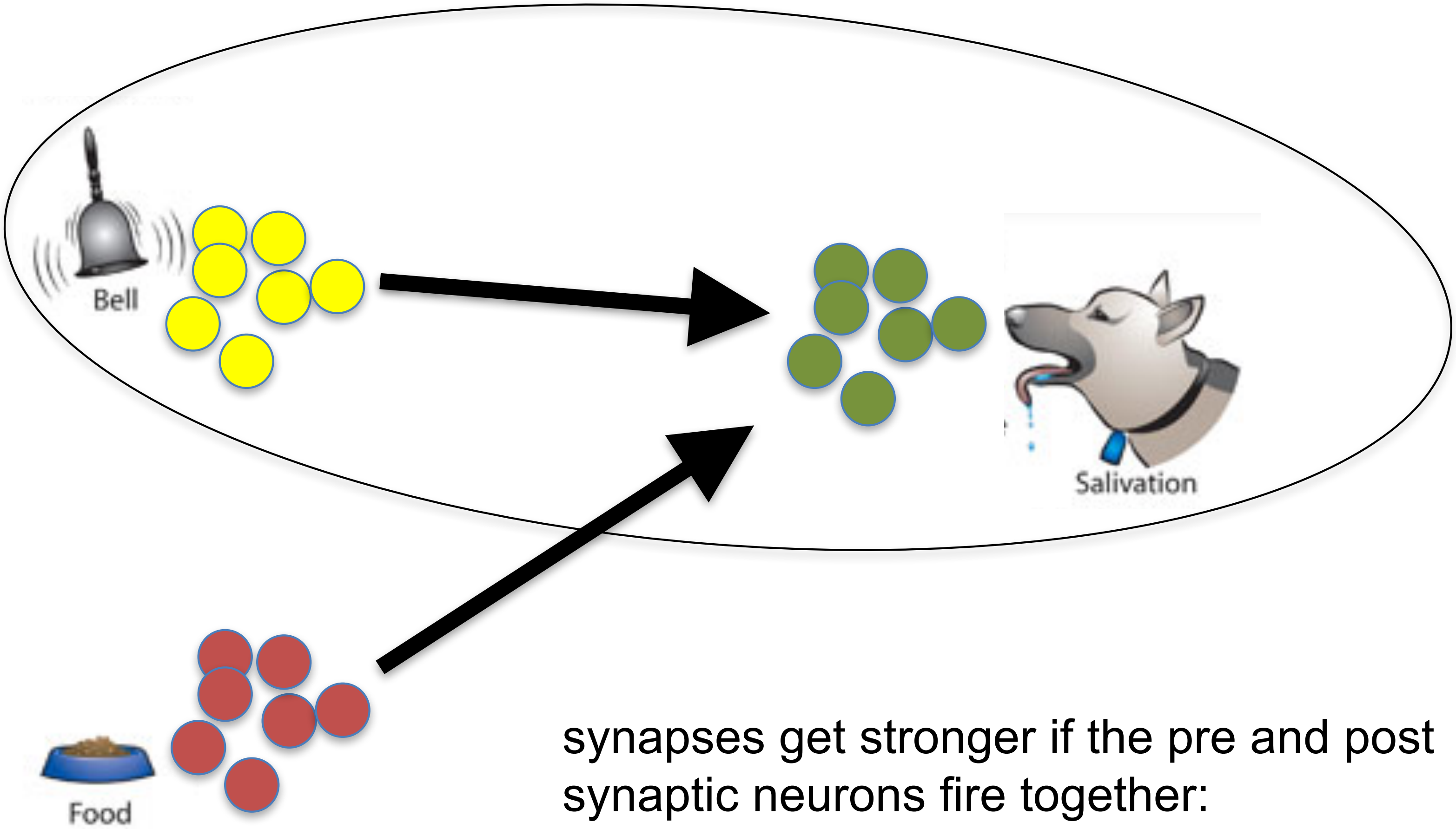
- What changes in a neural system when learning takes place?
- Consider a simple case of learning: Pavlov



Slide from Paul Munro (Pitt, SIS)

Hebb's Idea



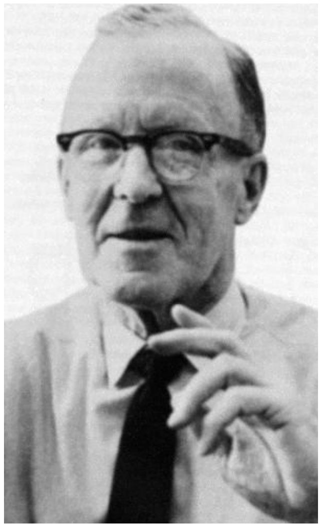


synapses get stronger if the pre and post synaptic neurons fire together:

*Neurons that **fire** together, **wire** together*

Slide from Paul Munro (Pitt, SIS)

Hebbian Learning



Hebb's Neurophysiological Postulate

"When an axon of cell *A* is near enough to excite cell *B* and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that *A*'s efficiency, as one of the cells firing *B*, is increased."

- D. O. Hebb (1949) *The Organization of Behavior*

Mathematical implementation (bilinear): $\Delta w_{ij}(t) = k\, a_i(t)\, a_j(t)$

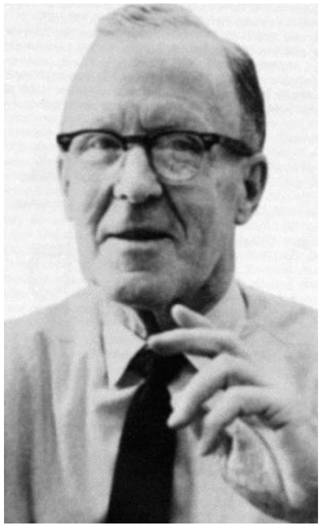
If $a_i(t)$ and $a_i(t)$ are never negative, the weights can never decrease by this rule -- they must either increase or not change... A mechanism for reducing weights is required.

Three approaches

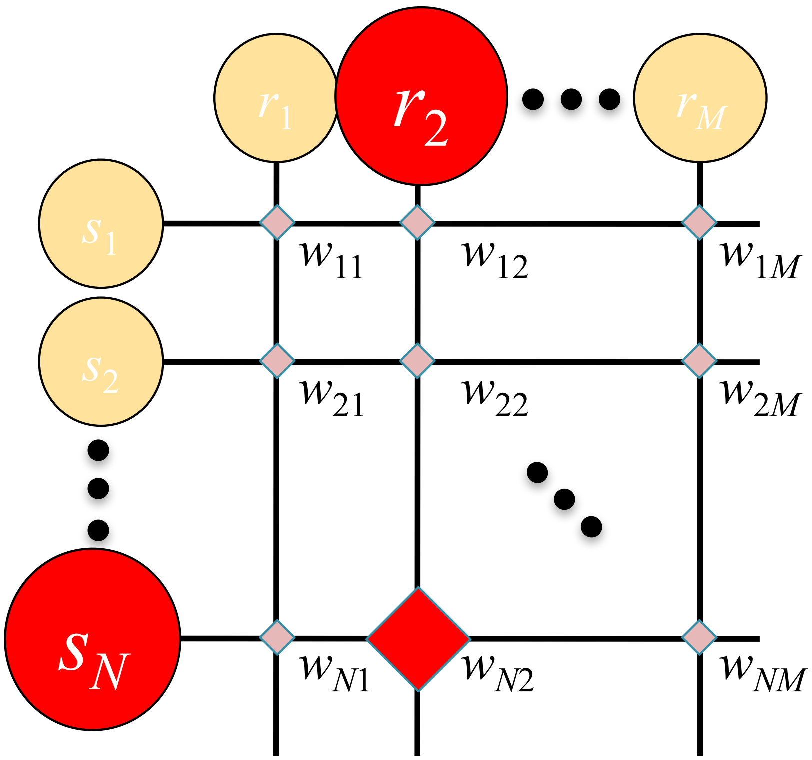
$\Delta w_{ij}(t) = \max\left[w_{ij}(t) + \Delta w_{ij}(t), U\right]$	**BSB**
$\Delta w_{ij}(t) = k\, a_i(t)\, a_j(t) - g w_{ij}(t)$	**Weight Decay**
$\Delta w_{ij}(t) = k\left[a_i(t) - \bar{a}_i\right]\left[a_j(t) - \bar{a}_j\right]$	**Subtract means**

Slide from Paul Munro (Pitt, SIS)

A matrix view of Hebbian Learning



Mathematical implementation (bilinear): $\Delta w_{ij}(t) = k\ a_i(t)\ a_j(t)$



Simplification:

$w_{ij}(0) = 0$

$k = 1$

$$w_{ij}(t) = \sum_{k=0}^{t} a_i(k)\ a_j(k)$$

Slide from Paul Munro (Pitt, SIS)

LINEAR ALGEBRA

Vector Spaces

Chalkboard

- Real-valued vectors of length n, R^n
- Properties of a vector space
- Inner product space
- Complete inner product space
- Other vector spaces

Vector Spaces

- **Definition:** A **field** is a set F of scalars (e.g. real numbers) that is closed under addition and multiplication and satisfies the following axioms
  1. addition and multiplication are **associative** and **commutative**
  2. there exists and **additive identity** and a **multiplicative identity** in F
  3. each element in F has an **additive inverse** and a **multiplicative inverse** (exception: the multiplicative identity does not have a multiplicative inverse)
  4. multiplication is **distributive** with respect to addition

- **Definition**: A **vector space** is a set V of *vectors* (e.g. real-valued vectors) that is closed addition (i.e. of vectors) and scalar multiplication over scalars from a field F and satisfies the following axioms
  1. addition and scalar multiplication are **associative** and **commutative**
  2. there exists an **additive identity** in V
  3. each vector has an **additive inverse** in V
  4. there exists a **multiplicative identity** in F
  5. each scalar has a **multiplicative inverse** in F
  6. scalar multiplication is **distributive** with respect to scalar addition and vector addition

Linear Algebra: Notation

Chalkboard

- Real-valued matrices
- Real-valued column/row vectors
- Elements of matrices/vectors
- Dot notation for columns/rows

Linear Algebra: Operations

Chalkboard

- Vector operations
  - dot product
  - outer product
- Matrix multiplication:
  vector-vector, matrix-vector, matrix-matrix
- Transpose
- Vector Norms
  - Euclidean norm
  - ℓ_p norms

HEBIAN LEARNING & MATRIX MEMORIES

Storing a Pattern Pair in a Matrix

In-Class Exercise

1. Given vectors:
   - stimulus: $x = [0.5, -0.5]^T$
   - response: $y = [1, 2]^T$
2. Compute their outer product:
   $W = yx^T$
3. Compute the predicted response:
   $r = Wx$
4. Does the relationship you observe here between r and y always hold?