Ling 8700: Lecture Notes 2
From Neural Activation to Associative Memory

The next few lectures will define a reference model to try to explain language:

1. We will model associative memory as relations between states of cortical activation.
2. We will model ideas as collections of cued associations in associative memory.
3. We will model language as a process of encoding, transmitting, and decoding ideas.

This first lecture is on associative memory.

Contents

2.1 Mental states as patterns of neural activation

Mental states (e.g. from looking at pictures) are associated with active firing of characteristic patterns of neurons (Mitchell et al., 2008).

Activation of neurons in the cortex can be modeled with **vectors** of firing rates for neurons or clusters:

.58	← neuron/cluster #1, say, closest to center of motor cortex
.0	← neuron/cluster #2, second closest to center of motor cortex
.58	← neuron/cluster #3, third closest to center of motor cortex
.0	⋮
.0	← neuron/cluster #5, say, closest to center of auditory cortex
.58	← neuron/cluster #6, second closest to center of auditory cortex
.0	⋮

(The values are typically 'normalized' so that the point is always one unit away from the origin.)

This kind of model is called 'distributed' because the activation is distributed around the cortex.

Individual elements of vectors (or characteristic subsets of elements) are called **features**.

An n-length vector may also be read as the **coordinates** of a point in an n-dimensional space.

2.2 Cued associations as connectivity weights between neurons/clusters

Mental states can be used as **cues** to other associated **target** mental states.

These associations happen by **long-term potentiation** (sensitization) of synapses between pre-synaptic and post-synaptic neurons that are active in the cue and target states, respectively.

This potentiation can be modeled using **matrices** of connections for each pair of neurons/clusters in cue and target patterns (specifically it's an outer product of cue and target vectors, with the cue on the right) (Marr, 1971; Anderson, Silverstein, Ritz, & Jones, 1977; Murdock, 1982; Smolensky, 1990; McClelland, McNaughton, & O'Reilly, 1995; Howard & Kahana, 2002):

$$\overbrace{\begin{bmatrix} .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \end{bmatrix}}^{\text{synaptic weights (cue:columns; target:rows)}} = \overbrace{\begin{bmatrix} .58 \\ .0 \\ .58 \\ .0 \\ .0 \\ .58 \\ .0 \end{bmatrix}}^{\text{target}} \overbrace{\begin{bmatrix} .0 & .50 & .50 & .0 & .50 & .0 & .50 \end{bmatrix}}^{\text{cue}}$$

This is just a matrix product: the value at row i, column j of the result is the sum of the product of each element in row i of the first factor with the corresponding element in column j of the second:

$$(FF')_{[i,j]} = \sum_k F_{[i,k]} F'_{[k,j]}$$

The target may then be obtained by applying the association weights to the cue (matrix product):

$$\overbrace{\begin{bmatrix} .58 \\ .0 \\ .58 \\ .0 \\ .0 \\ .58 \\ .0 \end{bmatrix}}^{\text{target}} = \overbrace{\begin{bmatrix} .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \end{bmatrix}}^{\text{synaptic weights (cue:columns; target:rows)}} \overbrace{\begin{bmatrix} .0 \\ .50 \\ .50 \\ .0 \\ .50 \\ .0 \\ .50 \end{bmatrix}}^{\text{cue}}$$

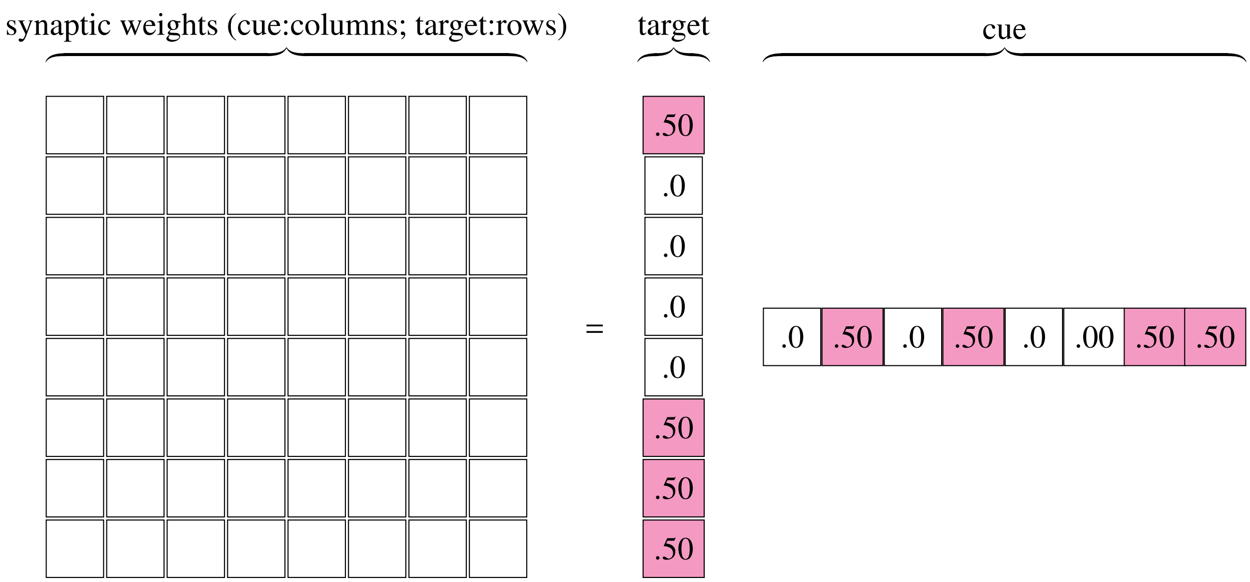
To compute the activation of each post-synaptic neuron (row) in the target vector, the activation of each of the pre-synaptic neurons in the cue is multiplied by the synaptic weight in the memory matrix for that pre-synaptic neuron (column) synapsing with that post-synaptic neuron (row). The contributions of each pre-synaptic neuron are weighted by the corresponding synaptic weight and added together. So, to compute the top element of the target, the four .50's of the 2nd, 3rd, 5th and 7th elements of the cue are multiplied by .29, .29, .29 and .29, respectively (the 2nd, 3rd, 5th and 7th elements of the top row) and added together to give .58, and the other elements in the top row of the matrix are multiplied by zeros in the cue so they don't change anything when they are added in. The same thing happens for each lower row of the matrix, to define each lower element of the target, until you have the result in the figure.

Cued association seems to be directional (North Korea → China, but China ↛ North Korea).

(This will show up later in our discussion of priming.)

2.3 Practice

Suppose you have cue and target mental states characterized by the below patterns of cortical activation. What synaptic weights result from long-term potentiation of the cue state immediately followed by the target state:



Now suppose you cue the below associative memory matrix of synaptic weights with the below vector of cortical activations. What will be the result?

$$
\overbrace{\begin{bmatrix} \\ \\ \\ \\ \\ \\ \\ \end{bmatrix}}^{\text{target}}
=
\overbrace{\begin{bmatrix}
.0 & .0 & .0 & .0 & .0 & .0 & .0 & .0 \\
.0 & .0 & .0 & .0 & .0 & .0 & .0 & .0 \\
.0 & .25 & .25 & .0 & .25 & .0 & .25 & .0 \\
.0 & .25 & .25 & .0 & .25 & .0 & .25 & .0 \\
.0 & .0 & .0 & .0 & .0 & .0 & .0 & .0 \\
.0 & .25 & .25 & .0 & .25 & .0 & .25 & .0 \\
.0 & .25 & .25 & .0 & .25 & .0 & .25 & .0 \\
.0 & .0 & .0 & .0 & .0 & .0 & .0 & .0
\end{bmatrix}}^{\text{synaptic weights (cue:columns; target:rows)}}
\overbrace{\begin{bmatrix} .0 \\ .50 \\ .50 \\ .0 \\ .50 \\ .0 \\ .50 \\ .0 \end{bmatrix}}^{\text{cue}}
$$

2.4 Robustness to incomplete cues ('holographic memory')

Associations from incomplete cues yield complete (but weaker) targets:

$$
\overbrace{\begin{bmatrix} .29 \\ .0 \\ .29 \\ .0 \\ .0 \\ .29 \\ .0 \end{bmatrix}}^{\text{target}}
=
\overbrace{\begin{bmatrix}
.0 & .29 & .29 & .0 & .29 & .0 & .29 \\
.0 & .0 & .0 & .0 & .0 & .0 & .0 \\
.0 & .29 & .29 & .0 & .29 & .0 & .29 \\
.0 & .0 & .0 & .0 & .0 & .0 & .0 \\
.0 & .0 & .0 & .0 & .0 & .0 & .0 \\
.0 & .29 & .29 & .0 & .29 & .0 & .29 \\
.0 & .0 & .0 & .0 & .0 & .0 & .0
\end{bmatrix}}^{\text{synaptic weights}}
\overbrace{\begin{matrix} \begin{bmatrix} .0 \\ .0 \\ .0 \\ .0 \\ .50 \\ .0 \\ .50 \end{bmatrix} & \begin{matrix} \\ \leftarrow \text{missing!} \\ \leftarrow \text{missing!} \\ \\ \\ \\ \\ \end{matrix} \end{matrix}}^{\text{cue}}
$$

To compute the activation of each post-synaptic neuron (row) in the target vector, the activation of each of the pre-synaptic neurons in the cue is multiplied by the synaptic weight in the memory matrix for that pre-synaptic neuron (column) synapsing with that post-synaptic neuron (row). The contributions of each pre-synaptic neuron are weighted by the corresponding synaptic weight and added together. So, to compute the top element of the target, the two .50's of the 5th and 7th elements of the cue are multiplied by .29 and .29 (the 5th and 7th elements of the top row) and added together to give .29, and the other elements in the top row of the matrix are multiplied by zeros in the cue so they don't change anything when they are added in. The same thing happens for each lower row of the matrix, to define each lower element of the target, until you have the result in the figure.

This provides a natural model of brain plasticity following trauma.

2.5 Associations can be combined

Multiple associations can be combined (stored together) in the same set of synapses:

$$\underbrace{\begin{bmatrix}.58\\.0\\.58\\.0\\.0\\.58\\.0\end{bmatrix}}_{\text{target 1}}\underbrace{\begin{bmatrix}.0 & .58 & .0 & .0 & .58 & .0 & .58\end{bmatrix}}_{\text{cue 1}} + \underbrace{\begin{bmatrix}.0\\.50\\.50\\.0\\.50\\.0\\.50\end{bmatrix}}_{\text{target 2}}\underbrace{\begin{bmatrix}.0 & .0 & .58 & .58 & .0 & .58 & .0\end{bmatrix}}_{\text{cue 2}}$$

$$= \begin{bmatrix}.0 & .33 & .0 & .0 & .33 & .0 & .33\\ .0 & .0 & .0 & .0 & .0 & .0 & .0\\ .0 & .33 & .0 & .0 & .33 & .0 & .33\\ .0 & .0 & .0 & .0 & .0 & .0 & .0\\ .0 & .0 & .0 & .0 & .0 & .0 & .0\\ .0 & .33 & .0 & .0 & .33 & .0 & .33\\ .0 & .0 & .0 & .0 & .0 & .0 & .0\end{bmatrix} + \begin{bmatrix}.0 & .0 & .0 & .0 & .0 & .0 & .0\\ .0 & .0 & .29 & .29 & .0 & .29 & .0\\ .0 & .0 & .29 & .29 & .0 & .29 & .0\\ .0 & .0 & .0 & .0 & .0 & .0 & .0\\ .0 & .0 & .29 & .29 & .0 & .29 & .0\\ .0 & .0 & .0 & .0 & .0 & .0 & .0\\ .0 & .0 & .29 & .29 & .0 & .29 & .0\end{bmatrix} = \begin{bmatrix}.0 & .33 & .0 & .0 & .33 & .0 & .33\\ .0 & .0 & .29 & .29 & .0 & .29 & .0\\ .0 & .33 & .29 & .29 & .33 & .29 & .33\\ .0 & .0 & .0 & .0 & .0 & .0 & .0\\ .0 & .0 & .29 & .29 & .0 & .29 & .0\\ .0 & .33 & .0 & .0 & .33 & .0 & .33\\ .0 & .0 & .29 & .29 & .0 & .29 & .0\end{bmatrix}$$

the resulting associations can still be cued:

$$\overbrace{\begin{bmatrix} .58 \\ .0 \\ .58 \\ .0 \\ .0 \\ .58 \\ .0 \end{bmatrix}}^{\text{target 1}} = \overbrace{\begin{bmatrix} .0 & .33 & .0 & .0 & .33 & .0 & .33 \\ .0 & .0 & .29 & .29 & .0 & .29 & .0 \\ .0 & .33 & .29 & .29 & .33 & .29 & .33 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .0 & .0 & .29 & .29 & .0 & .29 & .0 \\ .0 & .33 & .0 & .0 & .33 & .0 & .33 \\ .0 & .0 & .29 & .29 & .0 & .29 & .0 \end{bmatrix}}^{\text{synaptic weights}} \overbrace{\begin{bmatrix} .0 \\ .58 \\ .0 \\ .0 \\ .58 \\ .0 \\ .58 \end{bmatrix}}^{\text{cue 1}}$$

However, when the stored cues overlap (e.g. the 3rd element in the cues below):

$$\overbrace{\begin{bmatrix} .0 \\ .58 \\ .0 \\ .0 \\ .58 \\ .0 \\ .58 \end{bmatrix}}^{\text{target 1}} \overbrace{\begin{bmatrix} .58 & .0 & .58 & .0 & .0 & .58 & .0 \end{bmatrix}}^{\text{cue 1}} + \overbrace{\begin{bmatrix} .0 \\ .0 \\ .58 \\ .58 \\ .0 \\ .58 \\ .0 \end{bmatrix}}^{\text{target 2}} \overbrace{\begin{bmatrix} .0 & .50 & .50 & .0 & .50 & .0 & .50 \end{bmatrix}}^{\text{cue 2}}$$

$$= \begin{bmatrix} .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .33 & .0 & .33 & .0 & .0 & .33 & .0 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .33 & .0 & .33 & .0 & .0 & .33 & .0 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .33 & .0 & .33 & .0 & .0 & .33 & .0 \end{bmatrix} + \begin{bmatrix} .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .0 & .0 & .0 & .0 & .0 & .0 & .0 \end{bmatrix} = \begin{bmatrix} .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .33 & .0 & .33 & .0 & .0 & .33 & .0 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .33 & .0 & .33 & .0 & .0 & .33 & .0 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .33 & .0 & .33 & .0 & .0 & .33 & .0 \end{bmatrix}$$

the resulting associations ‘interfere’ with each other when cued, yielding a combined target:

$$\underbrace{\begin{bmatrix} .0 \\ .58 \\ .16 \\ .16 \\ .58 \\ .16 \\ .58 \end{bmatrix}}_{\text{combined target}} = \underbrace{\begin{bmatrix} .0 & .0 & .0 & .0 & .0 & .0 & .0 \\ .33 & .0 & .33 & .0 & .0 & .33 & .0 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .33 & .0 & .33 & .0 & .0 & .33 & .0 \\ .0 & .29 & .29 & .0 & .29 & .0 & .29 \\ .33 & .0 & .33 & .0 & .0 & .33 & .0 \end{bmatrix}}_{\text{synaptic weights}} \underbrace{\begin{bmatrix} .58 \\ .0 \\ .58 \\ .0 \\ .0 \\ .58 \\ .0 \end{bmatrix}}_{\text{cue 1}}$$

This has been proposed as a process by which forgetting happens (Howard & Kahana, 2002).

2.6 Graphical representations of mental states and cued associations

Recall mental states are coordinates of points in mental space, linked by cued associations:

$$\underbrace{\begin{bmatrix} .71 \\ .0 \\ .71 \\ .0 \\ .0 \end{bmatrix}}_{\text{target } v_1} \underbrace{\begin{bmatrix} .0 & .58 & .0 & .58 & .58 \end{bmatrix}}_{\text{cue } u_1} + \underbrace{\begin{bmatrix} .0 \\ .58 \\ .58 \\ .0 \\ .58 \end{bmatrix}}_{\text{target } v_2} \underbrace{\begin{bmatrix} .0 & .0 & .71 & .71 & .0 \end{bmatrix}}_{\text{cue } u_2} = \begin{bmatrix} .0 & .41 & .0 & .41 & .41 \\ .0 & .0 & .41 & .41 & .0 \\ .0 & .41 & .41 & .82 & .41 \\ .0 & .0 & .0 & .0 & .0 \\ .0 & .0 & .41 & .41 & .0 \end{bmatrix}$$

Cued associations in an associative memory can be represented graphically:

$$\begin{tikzcd} & u_1 \arrow[dl] & & u_2 \arrow[dl] \\ v_1 & & v_2 & \end{tikzcd}$$

(arbitrarily squashing the n coordinates/dimensions into a two-dimensional figure).

2.7 Multiple associations (multiplexing and tensors)

Initial part of hippocampus (dentate gyrus) is characterized as doing ‘pattern separation’ (Marr, 1971; Treves & Rolls, 1994; O’Reilly & McClelland, 1994; Hasselmo & Wyble, 1997):

- Input neurons multiply from .25M in entorhinal cortex (before hippocampus) to 1M in DG.
- Signals in DG observed in rats to be sparser and stronger.

Analogy to demultiplexer:

- Allows part of input cue (say first N elements) to act as 'switching' pattern.
- Transforms input into sparser set of pairwise/joint features.

Associations are then cued with this transformed representation, which produces less interference.

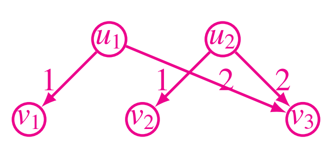
We can define a joint feature for each switch+non-switch element: on if both original elements on.

We can then model associations from joint features using numbered layered matrices (tensors):

$$\overbrace{\begin{bmatrix}.71\\.0\\.71\\.0\\.0\end{bmatrix}}^{\text{target } v_1}\overbrace{\begin{bmatrix}.0 & .58 & .0 & .58 & .58\end{bmatrix}}^{\text{cue } u_1}+\overbrace{\begin{bmatrix}.0\\.58\\.58\\.0\\.58\end{bmatrix}}^{\text{target } v_2}\overbrace{\begin{bmatrix}.0 & .0 & .71 & .71 & .0\end{bmatrix}}^{\text{cue } u_2}=\overbrace{\begin{bmatrix}.0 & .41 & .0 & .41 & .41\\ .0 & .0 & .41 & .41 & .0\\ .0 & .41 & .41 & .82 & .41\\ .0 & .0 & .0 & .0 & .0\\ .0 & .0 & .41 & .41 & .0\end{bmatrix}}^{\text{associations } M_1}$$

$$\overbrace{\begin{bmatrix}.58\\.58\\.0\\.58\\.0\end{bmatrix}}^{\text{target } v_3}\overbrace{\begin{bmatrix}.0 & .58 & .58 & .0 & .58\end{bmatrix}}^{\text{cue } u_1}+\overbrace{\begin{bmatrix}.58\\.58\\.0\\.58\\.0\end{bmatrix}}^{\text{target } v_3}\overbrace{\begin{bmatrix}.0 & .0 & .71 & .71 & .0\end{bmatrix}}^{\text{cue } u_2}=\overbrace{\begin{bmatrix}.0 & .33 & .74 & .41 & .33\\ .0 & .33 & .74 & .41 & .33\\ .0 & .0 & .0 & .0 & .0\\ .0 & .33 & .74 & .41 & .33\\ .0 & .0 & .0 & .0 & .0\end{bmatrix}}^{\text{associations } M_2}$$

Numbered association layers can be represented graphically using edge labels:



This allows mental states to cue multiple targets without interference, distinguished by assoc label.

References

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception and probability learning: Some applications of a neural model. *Psychological Review*, *84*, 413–451.

Hasselmo, M. E., & Wyble, B. P. (1997). Free recall and recognition in a network model of the hippocampus: simulating effects of scopolamine on human memory function. *Behavioural and Brain Research*, *89*, 1–34.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *45*, 269–299.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society (London) B*, *262*, 23–81.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*, 1191–1195.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*, 609–626.

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. *Hippocampus*, *4*, 661–682.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, *46*(1-2), 159–216.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*, 374–391.