1. Pros and Cons: Use Both.

2. Inhibition is also an Important Bias.

Biology Says: Both

Err Hebb Combo

\[ x_i, y_j \approx 0 \]

\[ x_i, y_j \approx 1 \]

\[ x_i - i, y_j - j \approx 0 \]

\[ x_i - i, y_j - j \approx 1 \]

No \[ \text{Ca}^{2+} \rightarrow \text{no learning} \]

Mod \[ \text{Ca}^{2+} \rightarrow \text{LTD} \]

High \[ \text{Ca}^{2+} \rightarrow \text{LTP} \]

\[ \Theta - \Theta \]

\[ \Theta + \Theta \]

\[ \Theta - \Theta \]

\[ \Theta + \Theta \]

\[ \text{LTD LTP} \]

\[ \text{Ca}^{2+} \]

\[ \text{only case disagreeing w/ Hebb is bottom left quadrant (LTD).} \]

\[ \text{Low Ca}^{2+} \text{ (just minus phase) produces LTD} \]

\[ \text{same mechanism as Hebb LTD for } y_i = 1, x_i = 0. \]

Error-driven Learning

• Neuromodulatory signals: Dopamine, Acetylcholine, etc.

• "Phasic" signals elicited by brain systems computing 'expected reward' and deviations from this expectation

• Resulting signals, when combined with target information (what should have been expected) in subsequent state, can enhance contrast between two succeeding attractor states

• Lots of evidence that LTP, LTD under neuromodulatory control

• Hebbian learning always occurs locally, in every synapse (model learning, statistics)

• Brain regions innervated by DA, ACh have enhanced weight changes during errors, leading to contrastive Hebbian learning (approximated by delta rule)

3. Other mechanisms for Error-driven Learning

Combining Error-driven + Hebbian

Get benefits of both:

\[ \Delta w_{ij} \approx \Delta_{\text{hebb}} + \Delta_{\text{err}} \] (1)

\[ \Delta_{\text{hebb}} = \epsilon_j a \]

\[ \Delta_{\text{err}} = \epsilon (a_i + a_j) - (a_i - a_j) \]

\[ \Delta w_{ij} = (k_{\text{hebb}}) \Delta_{\text{hebb}} + (1 - k_{\text{hebb}}) \Delta_{\text{err}} \] (2)

Hebbian bias helps so that weights are constrained to smaller set of solutions (otherwise too interdependent in err-driven)

Inhibitory Competition as a Bias

• Causes sparse, distributed representations (many alternatives, only a few relevant at any time).

• Competition and specialization, survival of fittest.

• Self-organizing learning.

• Often more important than Hebbian bias

Combining Error-driven + Hebbian

Error-driven = Left-wing, Hebbian = Right-wing (?!)

5. Combining Error-driven + Hebbian

Functional: Pros and Cons

<table>
<thead>
<tr>
<th></th>
<th>Error-driven</th>
<th>Hebbian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomous</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Reliable</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Greedy</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Myopic</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Remote</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Co-dependent</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Cooperative</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

2. Inhibition is also an Important Bias.

1. Hebb and Error both

Hebb says: Both

Combining Model & Task Learning
How do we do it?

Researcher's goal to learn

We are constantly faced with new situations and sensations

You hear, etc.

each time you walk into class, each social interaction, each sentence

how well do we deal with things we've never seen before?

Generalization

Distributed reps: novel items are novel combinations of existing features (combinatorial representations): "nust"

Input -> Comb. -> Internal reps

Hebbian & inhibition: produce elemental, combinatorial reps.

Bidirectional Activation Propagation

5. Error-driven Learning

4. Bidirectional

3. Inhibitory Competition

2. Distributed representations. Result?

1. Biological Realism

2. Distributed Representations

3. Inhibitory Competition

4. Bidirectional Activation Propagation

Generalization

How well do we deal with things we've never seen before?

The Whole Enchilada
Sims: [model and task proj.

Hebb:
• Sometimes fails to learn the training set
• Represents meaningful "things" in the world (correlations)
• Shows good generalization

Error (GeneRec):
• Always learns the training set
• Representations are "mushy"
• Can show poor generalization

Error + Hebb:
• Learns the training set (more quickly than error alone)
• Represents meaningful "things" in the world (correlations)
• Shows good generalization!

Deep Networks
Need many hidden layers to achieve many stages of transformation
But then the error signals are very remote & weak.
Need to add constraints and self-organizing learning:

- Constraints
- Flexibility
- Limits
- Inhib competition restricts flexibility (only certain states are valid)
- Combined model + task proj.
→ fewer degrees of freedom to adapt
Example: Family Trees (Hinton, 1986)

24 people, 12 relationships (brother, mother, granddaughter, etc.)

Who is Alf's grandmother?
Who is Lucia's daughter?