Übungen (Practicals)

- [http://www.physik3.gwdg.de/cns](http://www.physik3.gwdg.de/cns)

- 4 exercise sheets, over three weeks, each Covering contents parallel to the lectures

- Required: **C++ programming** and **gnuplot**
  i.e., lecture *Computational Physics*

- Submission of programs +
  10 min **presentation** and **discussion**

- **Details:** First session right after the lecture in the PC pool: **Room C.00.110**
Learning and Memory

Learning:
Learning Types/Classes and Learning Rules (Overview)
  Conceptualizing about Learning
  Math (Rules, Algorithms and Convergence) together with the
  Biological Substrate for different learning rules and
  Biological- and some other Applications (Pattern Recogn.,
  Robotics, etc.)

Memory:
Theories
Biological Substrate

Integrative Models – towards Cognition
Different Types/Classes of Learning

➢ Unsupervised Learning (non-evaluative feedback)
  • Trial and Error Learning.
  • No Error Signal.
  • No influence from a Teacher, Correlation evaluation only.

➢ Reinforcement Learning (evaluative feedback)
  • (Classic. & Instrumental) Conditioning, Reward-based Lng.
  • “Good-Bad” Error Signals.
  • Teacher defines what is good and what is bad.

➢ Supervised Learning (evaluative error-signal feedback)
  • Teaching, Coaching, Imitation Learning, Lng. from examples and more.
  • Rigorous Error Signals.
  • Direct influence from a Teacher/teaching signal.
Overview over different methods

Machine Learning
- Anticipatory Control of Actions and Prediction of Values
  - Reinforcement Learning
    - Example based
    - Dynamic Prog. (Bellman Eq.)
    - \( \delta \)-Rule
    - Eligibility Traces
      - TD(\(\lambda\))
        - often \(\lambda = 0\)
      - TD(1)
      - TD(0)
    - Monte Carlo Control
      - SARSA
      - Q-Learning
    - SARSA
  - Q-Learning
  - SARSA
  - Q-Learning
- Neuronal Reward Systems (Basal Ganglia)
  - ISO-Control
  - ISO-Control
  - ISO-Control

Classical Conditioning
- Anticipatory Control of Actions and Prediction of Values
  - Hebb-Rule
    - Rescorla/Wagner
    - Neur. TD-models
      - Neur. TD-formalism
      - Neur. TD-models
        - ("Critic")
    - Differential Hebb-Rule
      - ("slow")
    - ISO-Learning
      - ISO-Learning
      - ISO-Learning
    - Biophys. of Syn. Plasticity
      - Dopamine
      - Glutamate
    - STDP-Models
      - biophysical & network
      - STDP-Models
        - ISO-Model of STDP
  - STDP-Models
  - STDP-Models
  - STDP-Models

Synaptic Plasticity
- Correlation of Signals
  - Hebb-Rule
  - Differential Hebb-Rule
    - ("fast")
  - LTP (LTD = anti)
  - STDP-Models
  - STDP-Models
  - STDP-Models

Evaluative Feedback (Rewards)
- SARSA
- Correlation based Control
- (non-evaluative)
- ISO-Control
- ISO-Control
- ISO-Control

Non-Evaluative Feedback (Correlations)
- Neuronal Reward Systems (Basal Ganglia)
- ISO-Control
- ISO-Control
- ISO-Control

Biophys. of Syn. Plasticity
- Dopamine
- Glutamate
Overview over different methods

Machine Learning
- Reinforcement Learning: Example-based
  - Dynamic Prog. (Bellman Eq.)
  - Q-Learning
  - SARSA
  - Monte Carlo Control
  - Actor/Critic (Technical & Basal Ganglia)

Classical Conditioning
- Example-based
  - δ-Rule
  - Rescorla/Wagner

Eligibility Traces
- TD(λ)
  - Often λ = 0

ISO-Learning
- ISO-Models of STDP
- ISO-Control
- Correlation-based Control (Non-evaluative)
- Neuronal Reward Systems (Basal Ganglia)

Biophysical & Network
- Dopamine
- Glutamate
- STDP
- STDP-Models

Evaluation over different methods

Supervised Learning: Many more methods exist!
The Basics and a quick comparison
(before the maths really starts)

What can neurons compute?

What can networks compute?

Neurons can compute ONLY correlations!

Networks can compute anything 😊.

What is the biological Substrate for all learning?

The Synapse/synaptic strength (the connection strength between two neurons.)
The Neuroscience Basics as a Six Slide Crash Course

I forgot to make a backup of my brain.
All what I had learned last term is gone now.
Human Brain

Cortical Pyramidal Neuron
At the dendrite the incoming signals arrive (incoming currents).

At the soma current are finally integrated.

At the axon hillock action potential are generated if the potential crosses the membrane threshold.

The axon transmits (transports) the action potential to distant sites.

At the synapses are the outgoing signals transmitted onto the dendrites of the target neurons.

Structure of a Neuron:
Schematic Diagram of a Synapse

Terms to remember!
Ion channels consist of big (protein) molecules which are inserted into the membrane and connect intra- and extracellular space.

Channels act as a resistance against the free flow of ions: Electrical resistor $R$:

$$I_R = \frac{1}{R} (V_m - V_{rest}) = g (V_m - V_{rest})$$

If $V_m = V_{rest}$ (resting potential) there is no current flow. Electrical and chemical gradient are balanced (with opposite signs).

Channels are normally ion-selective and will open and close in dependence on the membrane potential (normal case) but also on (other) ions (e.g. NMDA channels).

Channels exists for: $K^+$, $Na^+$, $Ca^{2+}$, $Cl^-$
What happens at a chemical synapse during signal transmission:

The pre-synaptic action potential depolarises the axon terminals and Ca$^{2+}$-channels open.

Ca$^{2+}$ enters the pre-synaptic cell by which the transmitter vesicles are forced to open and release the transmitter.

Thereby the concentration of transmitter increases in the synaptic cleft and transmitter diffuses to the postsynaptic membrane.

Transmitter sensitive channels at the postsynaptic membrane open. Na$^+$ and Ca$^{2+}$ enter, K$^+$ leaves the cell. An excitatory postsynaptic current (EPSC) is thereby generated which leads to an excitatory postsynaptic potential (EPSP).
Information is stored in a Neural Network by the **Strength** of its Synaptic Connections

Up to 10000 Synapses per Neuron
Learning
Time Scales

Memory

- Working memory

- Short-term memory

- Long-term memory

Physiology

- Activity

- Short-term plasticity

- Long-term plasticity

- Structural plasticity

An unsupervised learning rule:

Basic Hebb-Rule: \[
\frac{d\omega_i}{dt} = \mu \, u_i \, v \quad \mu \ll 1
\]

For Learning: One input, one output

A reinforcement learning rule (TD-learning):

\[
\omega_{i,t+1} = \omega_{i,t} + \mu \left[ r_{t+1} + \gamma v_{t+1} - v_t \right] u_{i,t}
\]

One input, one output, one reward

A supervised learning rule (Delta Rule):

\[
\omega_{i,t+1} = \omega_{i,t} - \mu \frac{dE_t}{d\omega_i}
\]

No input, No output, one Error Function Derivative, where the error function compares input- with output-examples.
How can correlations be learned?

Correlation based (Hebbian) learning…

…correlates inputs with outputs by the…

…Basic Hebb-Rule: \( \frac{d\omega_1}{dt} = \mu u_1 v \) \( \mu \ll 1 \)

This rule is temporally symmetrical!
Conventional Hebbian Learning

\[ \frac{d\omega_1}{dt} = \mu u_1 v \]

Symmetrical Weight-change curve

The temporal order of input and output does not play any role
Our Standard Notation

\[ \frac{d\omega_1}{dt} = \mu u_1 \times v \]

Correlation between Input and Output

Hebbian Learning

Synapse = Amplifier with variable weight \( \omega_1 \)

Neuron (will sum different inputs, here only one)

Output

Input

\[ u_1 \]

\[ \omega_1 \]
Compare to Reinforcement Learning (RL)

This is Hebb!
What is this Eligibility Trace $E$ good for?
Classical Conditioning

I. Pavlov
What is this Eligibility Trace $E$ good for?

We start by making a single compartment model of a dog. The reductionist approach of a theoretician: The first stimulus needs to be “remembered” in the system.
TD Learning

Condition for convergence: \( \delta = 0 \)

\[
\delta^t = r^{t+1} + \gamma v^{t+1} - v^t
\]

Measured at the Output of the System
(Output Control)
Correlation based learning: No teacher
Reinforcement learning, indirect influence
Reinforcement learning, direct influence
Supervised Learning, Teacher
Programming
Open Loop versus Closed Loop Systems
The Difference between Input and Output Control

Input Control at the agent’s own sensors

True internal Value systems

Reinforcement

Output Control through observation of the agent

External Value systems

Evaluative Feedback (Klopf 1988)

Non-Evalulative Feedback
Is this a “real” or just an “academic” Problem: Why would we want Input Control?

The Output control paradigm can and does lead to major problems in reinforcement learning, because the wrong behaviour might be reinforced.
A “funny” example

Agent with own Goals (Me !)

Prior knowledge and shared goals help!

"Marsian" observer with other Sensors and/or other Goals

Bodyheat

Reinforces my running around

My perception of your attentiveness

Input Control Loop: Allows me to control my lecture and to learn to improve

Environment (You !)

Speech
Relevance for Learning:

1) Use output control to get a system that does what YOU want. (engineering system)

2) Use Input control to get an autonomous (biologically motivated system).

Other considerations:

Learning Speed

- Correlation based learning: No teacher
- Reinforcement learning, indirect influence
- Reinforcement learning, direct influence
- Supervised Learning, Teacher

Autonomy

Programming
• Good ending point
Is this a real or just an academic problem?

Observable Quantity

Agent that behaves (System)

What is the desired (most often occurring) output state?

Zero!

V

0
Experiment:

Assume you have one lever by which you can try to drive $V$ towards zero, whenever it suddenly deviates.
Obviously $V=0$ can be easily obtained when the lever follows $V$!
The System: A Braitenberg Vehicle

What the Agent wanted to learn was to approach the yellow food blob and eat it.
What you have reinforced:

\[ S_R = S_L = 0 \]

Leaving the food blob totally out of sight also gets \( V=0 \) (only the poor creature never eats and dies………)

The observable quantity \( V \) was not appropriate!
One should have observed \( A_R, A_L \) (but could not).

And……………. Things will get worse………. 🙃
Observer knows: “1) This is a Down-Neuron” (for Eye-Muscles)

Assumptions 1, 2 lead to Observer induced reinforcement

Observer knows: “2) There is evidence that the spatial ordering of synapses at a dendrite leads to direction selectivity and the observer has measured where the synapses are on the dendrite”

Assumptions 1 and 2 correspond to the Observer’s knowledge of this system
“This is a Down-Neuron” (for Eye-Muscles)

Observer

Synapses=States
Weights =Values

>1=reward

Environment

here also “Optics”

This Observer did lack the knowledge that the optics of the eye inverts the image

Motor Neuron (Actor)

Really this synapse should have been reinforced

Retinal receptive fields

True virtual image motion
A first order fallacy:

The observable quantity $V$ was not appropriate!
One should have observed $A_R, A_L$ (but could not).

A second order fallacy:

The observable quantities were appropriate but the Observer had a lack of knowledge about the inner signal processing in this system.
More realistically!

• Think of an engineer having to control the behavior and learning of a complex Mars-Rover which has many (1000?) simultaneous signals.
  – How would you know which signal configuration is at the moment beneficial for behavior and learning.

  → OUTPUT CONTROL WILL NOT WORK

• Ultimately only the Rover can know this.
  – But how would it maintain stability to begin with (not the be doomed from starters)
Since observers cannot have complete knowledge of the observed system we find that:

Output Control is fundamentally problematic.

A complex robot-world model required deep understanding on the side of the designer to define the appropriate reinforcement function(s).

This leads to a large degree of interference.

As a consequence the robot has then the world model of the designer (but not its own) – A slave not an autonomous agent.
\[ \Delta \rho_I = I \times II \times S > 0 \]
\[ \Delta \rho_{III} = III \times II \times S = 0 \]

Input Control will always work!

Environment here also “Optics”
The Chicken-Egg Problem Type I

Which came first: Chicken or Egg?
Control of my Input
(I, chook, want to feel an egg under my butt):

I, chook, would like to sit on this egg as long as required to hatch.

Control of its Output:

I, farmer, would like to get as many eggs as possible and take them away from the chook.

A fundamental Conflict

Autonomy  Servitude
Control from inside  Control from Outside
Value Systems (in the brain)

But that’s simple, isn’t it: Teaching will do it (supervised learning)!

You tell me, this is good and that is bad…………..

Supervised Learning

Reinforcement Learning – Learning from experience while acting in the world

I tell myself, this is good and that is bad…………..

a) Bootstrapping Problem: Evolution does not teach (evaluate).
b) Viewpoint Problem: Those are the values of the teacher and not of the creature.
c) Complexity Problem: SL requires already complex understanding.

Still: How do we get this in the first place?

Requires a Value-System in the Animal
(Dopaminergic System, Schultz 1998)
The Problem:
How to bootstrap a Value System?

Evolve it!

- Fully situated but takes long

Design it!

- Badly situated but can be achieved quickly