Reinforcement Learning
Intelligent Systems Series

Georg Martius

MPI for Intelligent Systems, Tübingen, Germany

October 19, 2018

Organizational structure of the lecture

- Two lecturers: me and Jie-Jie (JJ) Zhu
- Teaching language is English, although you can ask questions in German (to me)
- Fridays 12 c.t.–13:45 Lectures
- Fridays 14 c.t.–15:45 Recitations
- Exercises:
  - exercise sheets have to be returned in the following week
  - Need 50% passed sheets to be eligible for passing the course
  - Later in the course we will have projects
  - final exam will most likely be a presentation of the final project
- Lecture notes: There will be mixture between slides and black board, ... and background material to read
- Webpage: `http://al.is.tuebingen.mpg.de/pages/reinforcement-learning-ws-2018-19` (linked from Campus Verwaltung)
- On Dec 4th Lecture cannot take place, will find alternative time

Organizational structure of the lecture

- Two lecturers: me and Jie-Jie (JJ) Zhu
- Teaching language is English, although you can ask questions in German (to me)
- Fridays 12 c.t.–13:45 Lectures
- Fridays 14 c.t.–15:45 Recitations
- Exercises:
  - exercise sheets have to be returned in the following week
  - Need 50% passed sheets to be eligible for passing the course
  - Later in the course we will have projects
  - final exam will most likely be a presentation of the final project
- Lecture notes: There will be mixture between slides and black board, . . . and background material to read
- Webpage: `http://al.is.tuebingen.mpg.de/pages/reinforcement-learning-ws-2018-19` (linked from Campus Verwaltung)
- On Dec 4th Lecture cannot take place, will find alternative time

Organizational structure of the lecture

- Two lecturers: me and Jie-Jie (JJ) Zhu
- Teaching language is English, although you can ask questions in German (to me)
- Fridays 12 c.t.–13:45 Lectures
- Fridays 14 c.t.–15:45 Recitations
- Exercises:
  - exercise sheets have to be returned in the following week
  - Need 50% passed sheets to be eligible for passing the course
  - Later in the course we will have projects
  - final exam will most likely be a presentation of the final project
- Lecture notes: There will be mixture between slides and black board, …and background material to read
- Webpage: `http://al.is.tuebingen.mpg.de/pages/reinforcement-learning-ws-2018-19` (linked from Campus Verwaltung)
- On Dec 4th Lecture cannot take place, will find alternative time

Organizational structure of the lecture

- Two lecturers: me and Jie-Jie (JJ) Zhu
- Teaching language is English, although you can ask questions in German (to me)
- Fridays 12 c.t.–13:45 Lectures
- Fridays 14 c.t.–15:45 Recitations
- Exercises:
  - exercise sheets have to be returned in the following week
  - Need 50% passed sheets to be eligible for passing the course
  - Later in the course we will have projects
  - final exam will most likely be a presentation of the final project
- Lecture notes: There will be mixture between slides and black board, . . . and background material to read
- Webpage: `http://al.is.tuebingen.mpg.de/pages/reinforcement-learning-ws-2018-19` (linked from Campus Verwaltung)
- On Dec 4th Lecture cannot take place, will find alternative time

Organizational structure of the lecture

- Two lecturers: me and Jie-Jie (JJ) Zhu
- Teaching language is English, although you can ask questions in German (to me)
- Fridays 12 c.t.–13:45 Lectures
- Fridays 14 c.t.–15:45 Recitations
- Exercises:
  - exercise sheets have to be returned in the following week
  - Need 50% passed sheets to be eligible for passing the course
  - Later in the course we will have projects
  - final exam will most likely be a presentation of the final project
- Lecture notes: There will be mixture between slides and black board, ...and background material to read
- Webpage: `http://al.is.tuebingen.mpg.de/pages/reinforcement-learning-ws-2018-19` (linked from Campus Verwaltung)
- On Dec 4th Lecture cannot take place, will find alternative time

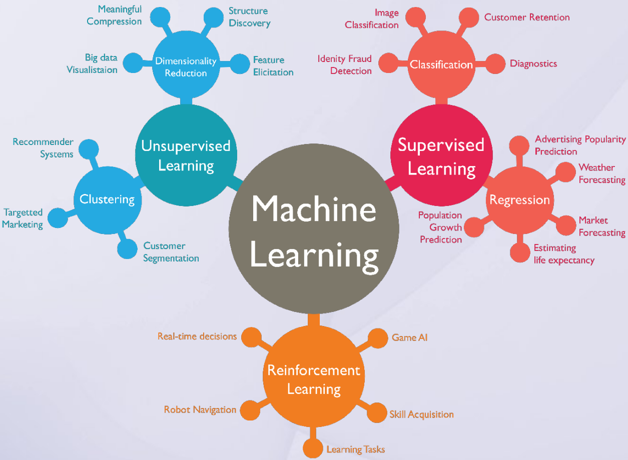
Organizational structure of the lecture

- Two lecturers: me and Jie-Jie (JJ) Zhu
- Teaching language is English, although you can ask questions in German (to me)
- Fridays 12 c.t.–13:45 Lectures
- Fridays 14 c.t.–15:45 Recitations
- Exercises:
  - exercise sheets have to be returned in the following week
  - Need 50% passed sheets to be eligible for passing the course
  - Later in the course we will have projects
  - final exam will most likely be a presentation of the final project
- Lecture notes: There will be mixture between slides and black board, ...and background material to read
- Webpage: `http://al.is.tuebingen.mpg.de/pages/reinforcement-learning-ws-2018-19` (linked from Campus Verwaltung)
- On Dec 4th Lecture cannot take place, will find alternative time

Machine Learning Overview

Machine learning is not voodoo,
it is about automatically finding a function that best solves a given task.

Three different classes of tasks:



Machine Learning Overview

Supervised Learning

given: $\{x, y\}_i \sim \mathcal{D}$ with data point $x \in \mathbb{R}^n$ and label $y \in \mathcal{Y}$ and $\mathcal{D}$ the data distribution.

What to find function $h(\cdot)$ such that

$$h(x) = y \quad \forall (x, y) \sim D$$

To measure quality of h and to be able to optimize something: Define loss function

$$J(h) = \mathbb{E}_{\mathcal{D}}[\text{dist}(y, h(x))]$$

(distance between true label y and predicted label $f(x)$)

Task: find function that minimized loss: $h^* = \arg\min_h J(h)$

Math can be so easy ;-)

We will see why this is not so easy in practice.

Machine Learning Overview

Supervised Learning

given: $\{x, y\}_i \sim \mathcal{D}$ with data point $x \in \mathbb{R}^n$ and label $y \in \mathcal{Y}$ and $\mathcal{D}$ the data distribution.
What to find function $h(\cdot)$ such that

$$h(x) = y \quad \forall (x, y) \sim D$$

To measure quality of h and to be able to optimize something: Define loss function

$$J(h) = \mathbb{E}_{\mathcal{D}}[\text{dist}(y, h(x))]$$

(distance between true label y and predicted label $f(x)$)

Task: find function that minimized loss: $h^* = \arg\min_h J(h)$

Math can be so easy ;-)

We will see why this is not so easy in practice.

Machine Learning Overview

Supervised Learning

given: $\{x, y\}_i \sim \mathcal{D}$ with data point $x \in \mathbb{R}^n$ and label $y \in \mathcal{Y}$ and $\mathcal{D}$ the data distribution.
What to find function $h(\cdot)$ such that

$$h(x) = y \quad \forall (x, y) \sim D$$

To measure quality of h and to be able to optimize something: Define loss function

$$J(h) = \mathbb{E}_{\mathcal{D}}[\text{dist}(y, h(x))]$$

(distance between true label y and predicted label $f(x)$)

Task: find function that minimized loss: $h^* = \arg\min_h J(h)$

Math can be so easy ;-)

We will see why this is not so easy in practice.

Machine Learning Overview

Supervised Learning

given: $\{x, y\}_i \sim \mathcal{D}$ with data point $x \in \mathbb{R}^n$ and label $y \in \mathcal{Y}$ and $\mathcal{D}$ the data distribution.
What to find function $h(\cdot)$ such that

$$h(x) = y \quad \forall (x, y) \sim D$$

To measure quality of h and to be able to optimize something: Define loss function

$$J(h) = \mathbb{E}_{\mathcal{D}}[\text{dist}(y, h(x))]$$

(distance between true label y and predicted label $f(x)$)

Task: find function that minimized loss: $h^* = \arg\min_h J(h)$

Math can be so easy ;-)

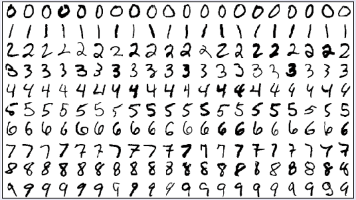
We will see why this is not so easy in practice.

Supervised Learning – Examples

Classification: $\mathcal{Y}$ is discrete

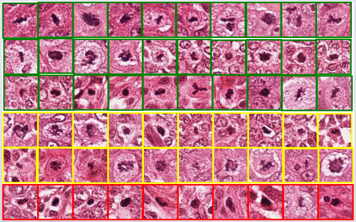
Examples:

Recognize handwritten digits:



(MNIST)

Classify pathology images:

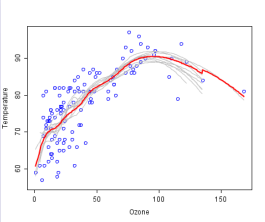


(Mitosis in breast cancer)

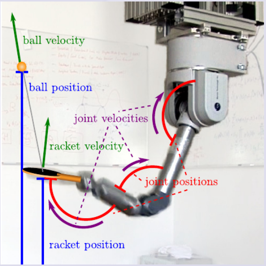
Regression: $\mathcal{Y}$ is continuous

Examples:

Predicting Ozon levels



Predicting torques



Machine Learning Overview

Unsupervised Learning

given: $\{x\}_i$ with $x \in \mathbb{R}^n$

What to find function $f(\cdot)$ such that $f(x) = y$ where y low dimensional,
e.g. a cluster number

- Much less clear what is the objective.
- Many algorithms but no unifying theory.

Machine Learning Overview

Unsupervised Learning

given: $\{x\}_i$ with $x \in \mathbb{R}^n$
What to find function $f(\cdot)$ such that $f(x) = y$ where y low dimensional,
e.g. a cluster number

- Much less clear what is the objective.
- Many algorithms but no unifying theory.

Machine Learning Overview

Unsupervised Learning

given: $\{x\}_i$ with $x \in \mathbb{R}^n$
What to find function $f(\cdot)$ such that $f(x) = y$ where y low dimensional,
e.g. a cluster number

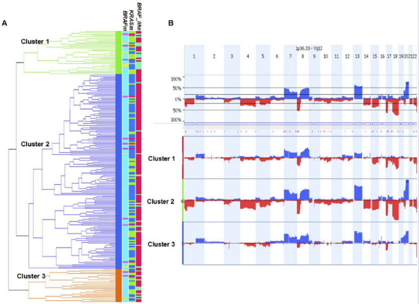
- Much less clear what is the objective.
- Many algorithms but no unifying theory.

Unsupervised Learning – Examples

Clustering: discrete y

Examples:

Genome comparison:



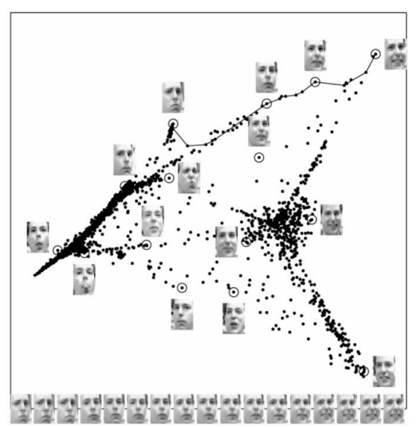
(by Tao Xie)

Both cases are expecially useful for high-dimensional data

Dim. reduction: continuous y

Examples:

Finding descriptors for face expressions



(by Sam T Roweis)

Reinforcement Learning

given:

- system to interact with: $s_{t+1} = S(a_t, s_t)$ where s_t is the state and a_t is the action.
- reward/utility function: $r_t = U(a_t, s_t)$

What to find function $f(\cdot)$ (policy) such that $a = f(s)$ and $\mathbb{E}[r]$ is maximized.

In general: stochastic systems formulated as Markov Decision Processes.

- Need to simultaneously learn f and potentially models of S and U.
- Reward can be sparse (e.g. only at the end of an long action sequence)

Notation: Later the policy will be denoted as $\pi(s)$

Machine Learning Overview

Reinforcement Learning

given:

- system to interact with: $s_{t+1} = S(a_t, s_t)$ where s_t is the state and a_t is the action.
- reward/utility function: $r_t = U(a_t, s_t)$

What to find function $f(\cdot)$ (policy) such that $a = f(s)$ and $\mathbb{E}[r]$ is maximized.

In general: stochastic systems formulated as Markov Decision Processes.

- Need to simultaneously learn f and potentially models of S and U.
- Reward can be sparse (e.g. only at the end of an long action sequence)

Notation: Later the policy will be denoted as $\pi(s)$

Machine Learning Overview

Reinforcement Learning

given:

- system to interact with: $s_{t+1} = S(a_t, s_t)$ where s_t is the state and a_t is the action.
- reward/utility function: $r_t = U(a_t, s_t)$

What to find function $f(\cdot)$ (policy) such that $a = f(s)$ and $\mathbb{E}[r]$ is maximized.

In general: stochastic systems formulated as Markov Decision Processes.

- Need to simultaneously learn f and potentially models of S and U.
- Reward can be sparse (e.g. only at the end of an long action sequence)

Notation: Later the policy will be denoted as $\pi(s)$

Machine Learning Overview

Reinforcement Learning

given:

- system to interact with: $s_{t+1} = S(a_t, s_t)$ where s_t is the state and a_t is the action.
- reward/utility function: $r_t = U(a_t, s_t)$

What to find function $f(\cdot)$ (policy) such that $a = f(s)$ and $\mathbb{E}[r]$ is maximized.

In general: stochastic systems formulated as Markov Decision Processes.

- Need to simultaneously learn f and potentially models of S and U.
- Reward can be sparse (e.g. only at the end of an long action sequence)

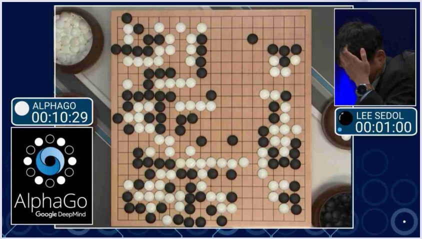
Notation: Later the policy will be denoted as $\pi(s)$

Reinforcement Learning – Examples

Robot Control

(by MPI-IS)

Deepmind AlphaGo



(go-baduk-weiqi.de)

Improve performance by learning from experience
and exploring new strategies.

Rough plan of the course

- Introduction to Supervised learning and Imitation learning
  - linear regression, regularization, model selection, ...
  - neural networks
  - behavior cloning, Data Aggregation...
- Reinforcement Learning
  - Markov Decision Processes (MDPs) and background
  - Bellman equations and TD learning, Q-Learning, ...
  - Deep Reinforcement Learning
  - Continuous Spaces:
    - Actor-Critic
    - Reinforcement Learning with parametrized policies
    - Episodic RL as parametrized optimization problem
  - Optimal control/trajectory optimization and model predictive control
  - Exploration vs exploitation
  - Advanced topics: goal-based RL, intrinsic motivation etc.