

# P01 – Reading Assignment

## 1. Background

Machine Learning is a very broad interdisciplinary field with contributions from artificial intelligence, statistics, theoretical computer science and other disciplines. It is also a very active and rapidly developing area of research. Hence, a wide body of different approaches and methods exist. That is why it is hard to cover all material in one course, and why it will always be necessary in the future for you to complement your knowledge with recent or additional inputs from the literature (scientific papers, text books, blog posts or instructional videos).

Thus, we bring you in contact with relevant literature right from the start. On the course site, you find the following references (among others):

Tag	Authors	Title	Year
<i>Murphy</i>	Murphy	Machine Learning - A Probabilistic Perspective	2012
<i>Mitchell</i>	Mitchell	Machine Learning	1997
<i>AIMA</i>	Russell, Norvig	Artificial Intelligence - A Modern Approach, 3rd Edition	2010

Accomplish the following reading assignment tasks 2 – 3 in the given order. Take brief but distinct notes on

- what is interesting to you
- seems relevant for the future
- is not understandable at the moment

...while you read.

To summarize your notes, write down at least three questions and two facts in task 5. Then, update (or get) your Anaconda Python installation to be ready for the next labs.

## 2. Chapters 1.2 – 1.4 from *[Murphy]*

Notes:

Interessant: Google SmartASS: Kannte den Begriff vorher noch nicht

---

Relevant: Unterteilung Supervised/Unsupervised Learning

---

---

Relevant: Classification und Regression (Supervised Learning), Clustering (Unsupervised Learning)

---

---

Relevant: PCA, Collaborative Filtering, K-Means, Overfitting

---

---

Relevant: Linear/Logistic Regression, Cross-Validation, Model Selection (no free lunch)

---

### 3. Chapters 1.1 – 1.2 from *[Mitchell]*

Notes:

---

Interessant: Viele aktuelle Themen (Spracherkennung, pers. Assistent, autonomes Fahren) waren schon 1997 aktuell

---

---

Relevant: Viele verschiedene Disziplinen beeinflussen ML

---

---

Relevant: Direct/indirect feedback, learning with/without teacher

---

---

Relevant: Approximation vs. real target function

---

---

Interessant: Beispiel Schachcomputer mit vielen Zwischen-Zuständen, welche nicht trivial sind

---

---

Relevant: Darstellung eines ML-Systems mit Performance System, Critic, Generalizer und Experiment Generator

---

---

Relevant: Loss Function, Regularization

---

## 4. Chapter 18.4 from [AIMA]

Notes:

Relevant: LOOCV, stratifizierte CV

---

Relevant: Regularization, Complexity vs. Loss

---

nicht ganz verstanden: Generalization Loss

---

Frage: Welche Methoden für Feature Selection gibt es?

---

## 5. Summarizing Facts & Questions

Summarize your reading in at least 2 facts (anything you find interesting/profound) and 3 questions (something you didn't understand or where you have follow-up questions):

Fact 1:

Unterteilung in Training-, Validation- und Test-Set ist entscheidend

---

Fact 2:

Manchmal sind einfache Modelle überraschend gut

---

Question 1:

---

Model Selection: Wie ist das effizient machbar? Gibt es Strategien/Faustregeln, welche trotz fehlender Erfahrung bei der Auswahl besserer und schlechterer Modelle helfen?

---

Question 2:

---

Welche Algorithmen/Strategien/Möglichkeiten/... für Feature Selection gibt es und wann werden welche angewendet? Wie funktioniert Chi-Squared-Pruning?

---

Question 3:

## 6. Installing and Using Anaconda

Continue with exercise 7 if you already have Anaconda installed. Other Python distributions may be used, but are not supported.

### 6.1. Background

Anaconda is a scientific Python distribution by Continuum Analytics. Anaconda offers:

- Easy installers for Windows, MacOS & Linux with all necessary libraries included
- A sophisticated but clear integrated development environment called “Spyder”

### 6.2. Downloading Anaconda

- Download the correct version for your system from here (ca. 515 MB):  
<http://continuum.io/downloads#all>
  - Be sure to use the Version for Python 3.x, *not* Python 2.x
  - Take the 64-bit variant if you have a 64 bit system

### 6.3. Installing Anaconda

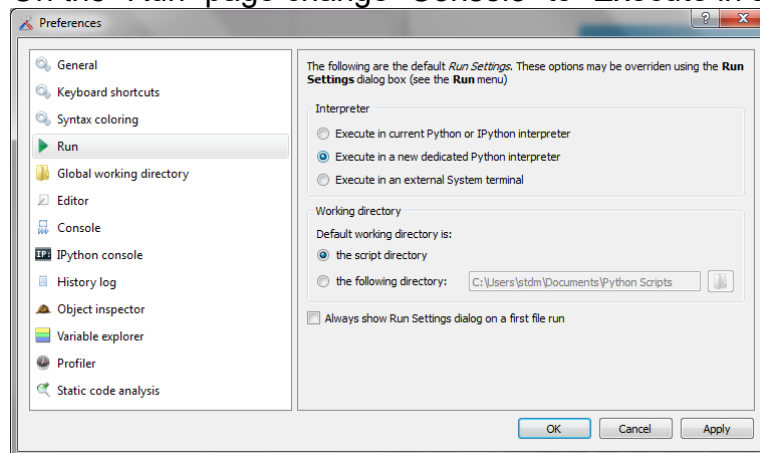
- Follow the instructions in the setup process
  - You can safely accept all standard choices
  - On Windows: If you are asked if you want to install for everyone or just yourself → choose “for everyone”

- On Windows: If you are asked if Anaconda should be your standard Python environment → conform with “yes”
- On MacOS X: if the installer says something like ‘cannot install’, choose ‘just install for me’.

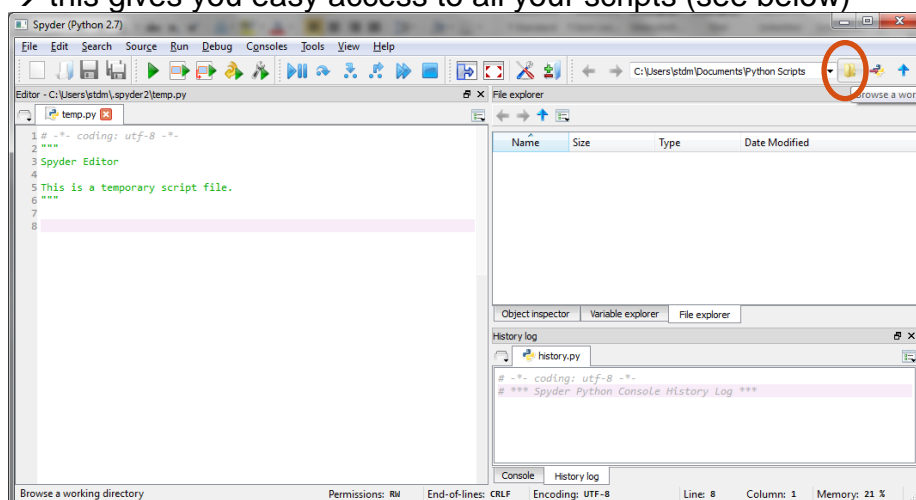
## 6.4. Configuring Spyder

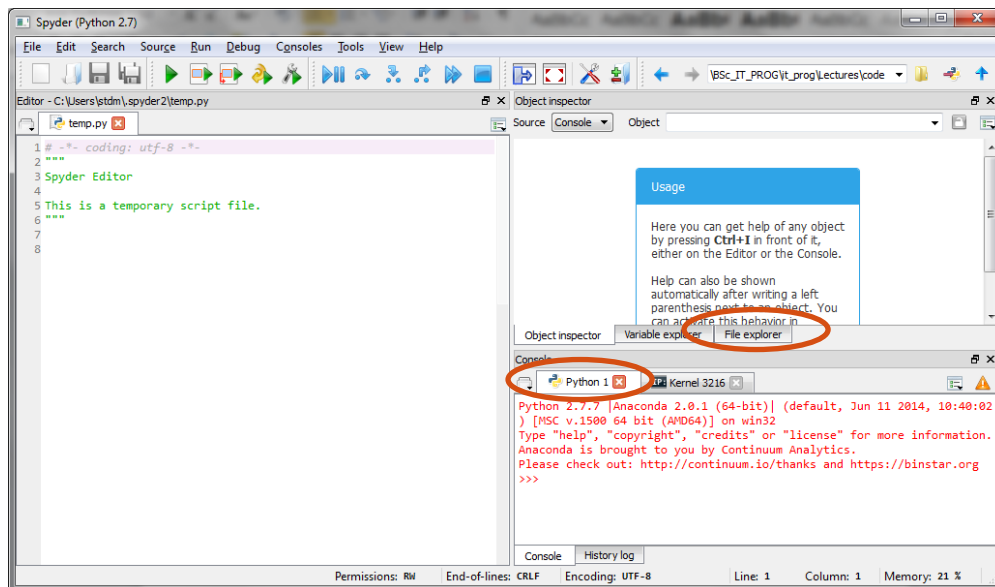
The following settings make Spyder look tidier:

- Start Spyder (on Mac: go to your home/anaconda and click on Launcher, then start Spyder)
- Go to “Tools” → “Preferences”
  - On the “Run” page change “Console” to “Execute in a dedicated console”



- Close the Preferences with “ok”
- Click the “Browse a working directory” Button at the right end of the menu bar and choose the folder where you store your scripts  
→ this gives you easy access to all your scripts (see below)

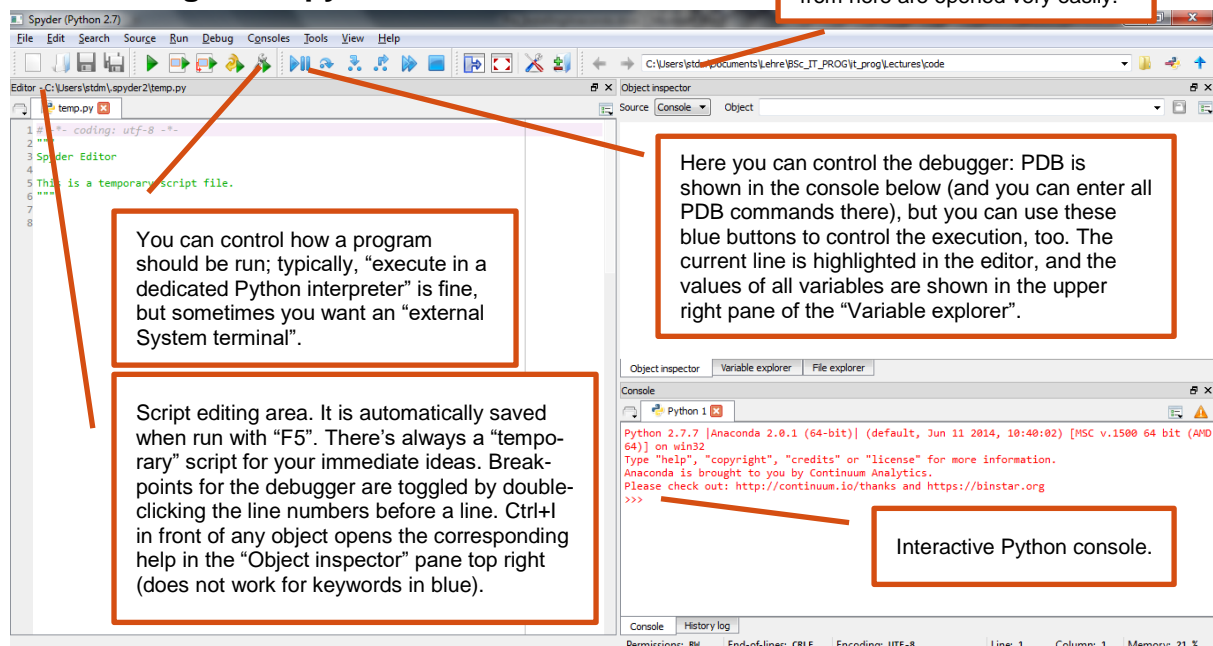




- In middle of the right half of the window, switch to the “File explorer” tab  
→ you now see a list of all the scripts in the current working directory and can easily open them

Close Spyder and start it again: All your settings should be stored and recovered.

## 6.5. Working with Spyder



- Go to the interactive console tab (bottom right) and use it as a calculator
- Go to temp.py in the code editor (left) and write your individual “hello world” script. Run it, modify it, re-run it.



## 7. Update your Anaconda Python installation

Anaconda (Python) needs to be up to date to be used for the labs accompanying this module. Please update it using either the Anaconda Launcher or the following two console command (from Anaconda install directory):

```
conda update conda  
conda update anaconda
```

See also: <http://docs.continuum.io/anaconda/install#updating-from-older-anaconda-versions>