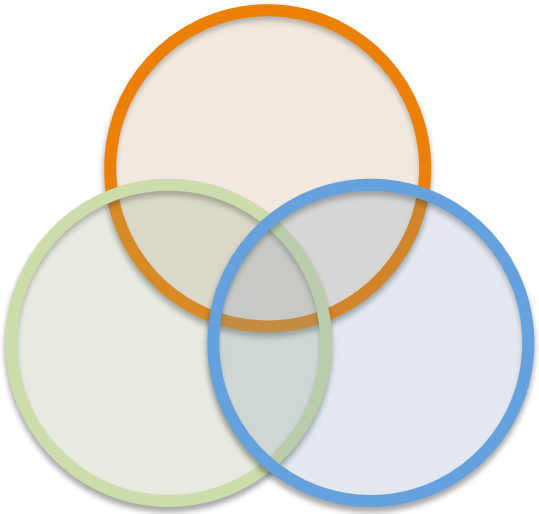
Machine Learning for the Quantified Self



Lecture 1
Introduction

Quantified Self definition (1)

- Term first coined by Gary Wolf and Kevin Kelly in Wired Magazine
- Let's watch a video first:

  https://www.youtube.com/watch?v=OrAo8oBBFIo

Quantified Self definition (2)

- What is the quantified self?
  - Swan (2013): "*The quantified self is any individual engaged in the self-tracking of any kind of biological, physical, behavioral, or environmental information. There is a proactive stance toward obtaining information and acting on it.*"

Quantified Self definition (3)

– We: "*The quantified self is any individual engaged in the self-tracking of any kind of biological, physical, behavioral, or environmental information. The self-tracking is driven by a certain goal of the individual with a desire to act upon the collected information.*"

Quantified Self: measurements

- Augemberg (2012):

Type of measurement	Examples
Physical activities	miles, steps, calories, repetitions, sets, METs (metabolic equivalents)
Diet	calories consumed, carbs, fat, protein, specific ingredients, glycemic index, satiety, portions, supplement doses, tastiness, cost, location
Psychological states and traits	mood, happiness, irritation, emotions, anxiety, self-esteem, depression, confidence
Mental and cognitive states and traits	IQ, alertness, focus, selective/sustained/divided attention, reaction, memory, verbal fluency, patience, creativity, reasoning, psychomotor vigilance
Environmental variables	location, architecture, weather, noise, pollution, clutter, light, season
Situational variables	context, situation, gratification of situation, time of day, day of week
Social variables	influence, trust, charisma, karma, current role/status in the group or social network

Quantified Self: why? (1)

- Choe, 2014:
  - Interview with 52 quantified selfs
  - Three categories:
    - Improved health (cure or manage a condition, execute a treatment plan, achieve a goal)
    - Improve other aspects of life (maximize work performance, be mindful)
    - Find new life experiences (have fun, learn new things)

Quantified Self: why? (2)

- Gimpel, 2013:
  - Identify “Five-Factor Framework of Self-Tracking Motivations:
    - Self-healing (become healthy)
    - Self-discipline (rewarding aspects of it)
    - Self-design (control and optimize “yourself”)
    - Self-association (associated with movement)
    - Self-entertainment (entertainment value)

Quantified Self: Arnold and Bruce

- Use two running examples

Arnold:

- Loves sports
- Wants to participate in IRONMAN
- Gadget freak
- Smart phone/watch/…
- Electronic scale
- Chest strap
- ……

Bruce:

- Diabetic
- Susceptible for depression
- Smart watch
- Device to measure blood glucose level
- ……

Moving on the machine learning

- Machine learning: “Machine learning is to automatically identify patterns from data”
- What could we learn for Arnold and Bruce?

What could we learn?

- Arnold:
  - Advising the training to make most progress towards a certain goal based on past outcomes of training.
  - Forecasting when a certain running distance will be feasible based on the progress made so far and the training schedule.
- Bruce:
  - Predict the next blood glucose level based on past measurements and activity levels.
  - Determine when and how to intervene when the mood is going down to avoid a spell of depression.
  - Finding clusters of locations that appear to elevate one's mood.

Why is the Quantified Self so different?

- Sensory noise
- Missing measurements
- Temporal data
- Interaction with a user
- Learn over multiple datasets

Basic Terminology (1)

- A measurement is one value for an attribute recorded at a specific time point.

Time point	The time point at which the measurement took place (considered in hours for this example)
Heart rate	Beats per minute, integer value
Activity level	Can be either low, medium or high
Speed	Speed in kilometers per hour, real value
Facebook post	A string representing the Facebook message posted
Activity type	The type of activity: inactive, walking, running, cycling, gym

Basic Terminology (2)

- A time series is a series of measurements in temporal order.

Time point	Heart rate	Activity level	Speed	Facebook post	Activity type
14:30	55	low	0	getting ready to hit the gym	inactive
14:45	55	low	0	having trouble getting off the couch	inactive
15:00	70	medium	5	walking to the gym, it's gonna be a great workout, I feel it	walking
15:10	130	high	0	-	gym
15:50	120	high	12	the gym didn't do it for me, running home	running
16:15	130	high	35	still have energy, on my bike now	cycling

Basic Terminology (3)

- Machine learning terminology is assumed to be known, for your convenience:
  - **Supervised learning** is the machine learning task of inferring a function from a set of labeled training data
  - In **unsupervised learning**, there is no target measure (or label), and the goal is to describe the associations and patterns among the attributes
  - **Reinforcement learning** tries to find optimal actions in a given situation so as to maximize a numerical reward that does not immediately come with the action but later in time.

Mathematical notation (1)

Notation	Explanation
	Dataset representation
X_k	A variable (or attribute) in our dataset, k is the index of the variable.
$\mathbf{X_i^{\mathcal{T}}}$	Matrix representing a dataset containing N_i instances with p variables. The i allows us to refer to a specific dataset (e.g. of a specific person) while the $^{\mathcal{T}}$ indicates a dataset with a temporal ordering. If $^{\mathcal{T}}$ is omitted no assumption about the ordering within the dataset is made.
x_j^k	The j^{th} observation in the dataset. k refers to the specific variable within the observation. If k is omitted this concerns an observation of the entire vector of variables.
	Categorical target representation (optional)
G	A categorical target variable in our dataset.
$\mathbf{G}$	Similar to $\mathbf{X_i^{\mathcal{T}}}$ (and the same additional super- and subscripts can be used), except that this refers to the categorical targets for our dataset (if present). It contains N_i instances.
g_j	The j^{th} row of categorical targets in $\mathbf{G}$.
	Classifier prediction representation
$\hat{g}_j$	The prediction of our classifier of the target for the j^{th} row in the dataset.
$\hat{\mathbf{G}}$	The entire set of categorical predictions of our classifier.
	Numerical target representation (optional)
Y	A numerical target variable in our dataset.
$\mathbf{Y}$	Similar to $\mathbf{X_i^{\mathcal{T}}}$ (and again the same additional super- and subscripts can be used), except that this refers to the numerical targets for our dataset (if present). It contains N_i instances.
y_j	The j^{th} row of numerical targets in $\mathbf{Y}$.
	Numerical prediction representation
$\hat{y}_j$	The prediction of our model of the numerical targets for the j^{th} row in the dataset.
$\hat{\mathbf{Y}}$	The entire set of numerical predictions of our model.

Mathematical notation (2)

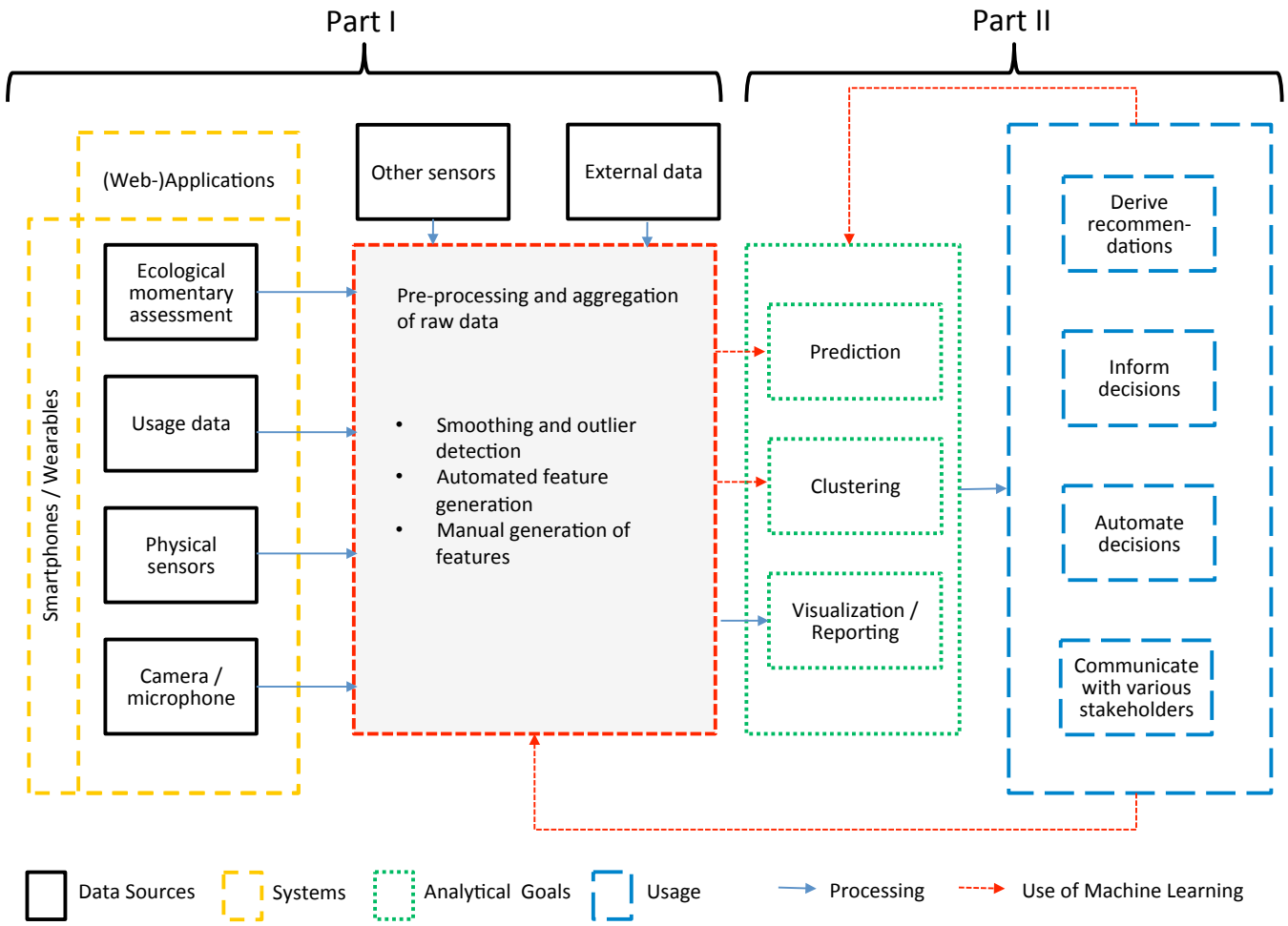
- Some examples:
  - An instance:

$$x_1 = \begin{bmatrix} 0 \\ 45 \\ \text{low} \\ 0 \\ \text{"getting ready to hit the gym"} \end{bmatrix}$$

  - A target for the instance:

$$g_1 = [inactive]$$

Overview of the course (1)



Overview of the course / the book (2)

- We will discuss the following topics:
  1. Sensory data (+ case study) (Chapter 2)
  2. Handling noise (Chapter 3)
  3. Feature extraction (Chapter 4)
  4. Clustering (Chapter 5)
  5. Theoretical foundations (Chapter 6)
     Predictive modeling without time (Chapter 7)
  6. Predictive modeling with time (Chapter 8)
  7. Reinforcement learning (Chapter 9)
     Look towards the future (Chapter 10)