Python Data Products

Course 1: Basics

Lecture: Our Case Study (Recommender Systems)

Learning objectives

In this lecture we will...
- Introduce the concept of Recommender Systems, which we'll use as an ongoing case-study throughout the Specialization
- Describe some of the main datasets we will use to study recommender systems and their main characteristics
- Motivate the use of these datasets for various problems throughout the Specialization

Recommender Systems

As a running example throughout this course, we will build **recommender systems** that model interactions between users and items

- In particular, we will focus on two publicly-available datasets from Amazon and Yelp:
- https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt
- https://www.yelp.com/dataset/download

Recommender Systems
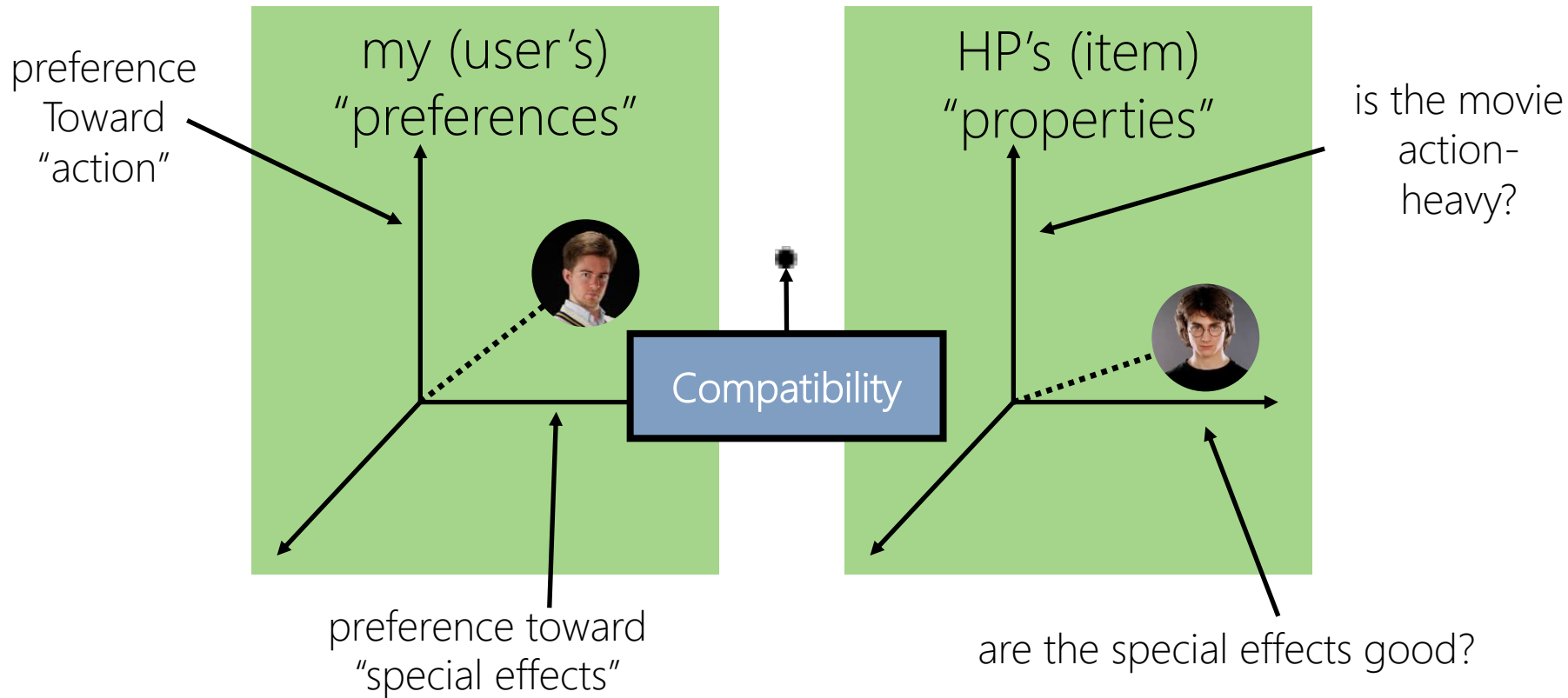
We are particularly interested in tasks including:

- How can we read and process these large datasets, containing complex, structured fields (**Course 1**)?
- How can we make simple predictions from these datasets, such as the sentiment of a review or the category of a business (**Course 2**)?
- How can we validate these predictions, and compare different modeling approaches (**Course 3**)?
- How can we build and deploy a working system using these predictive models (**Course 4**)?

What do recommender systems do?

(preference modeling)

(pricing)

existing items

(retrieval)
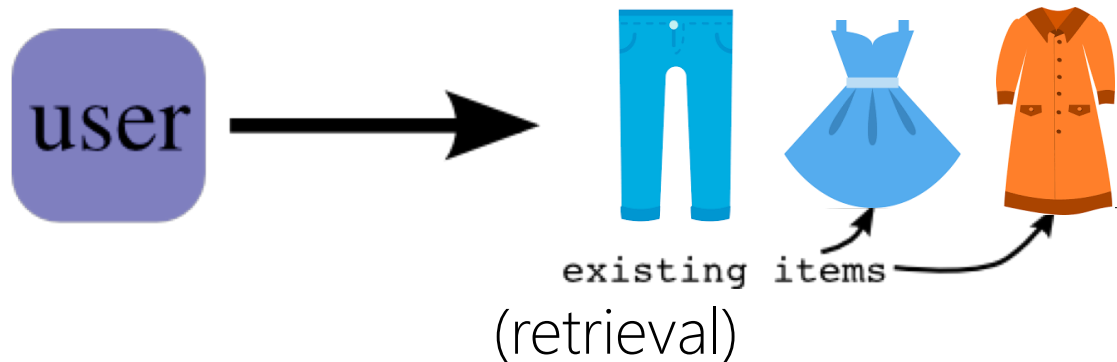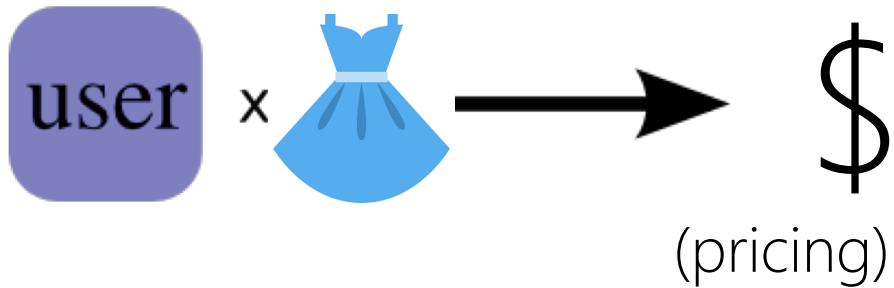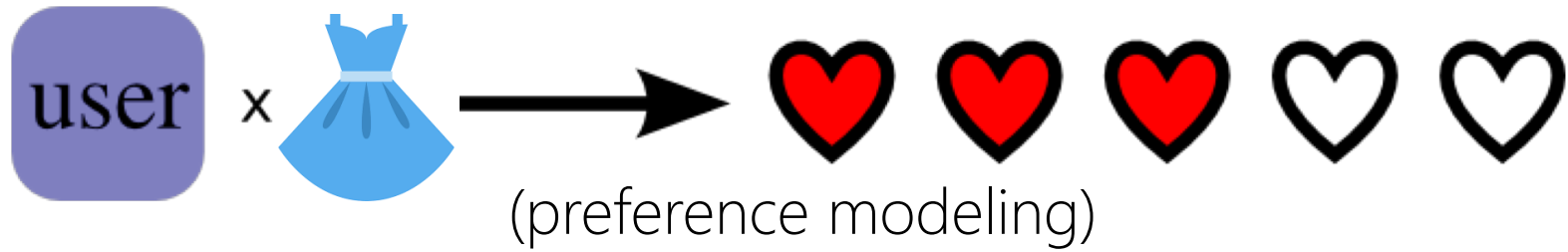
What do recommender systems do?

In essence, Recommender Systems work by trying to model the **relationships** between people and the items they're evaluating:

preference
Toward
"action"

my (user's)
"preferences"

HP's (item)
"properties"

is the movie
action-
heavy?

Compatibility

preference toward
"special effects"

are the special effects good?

Recommender Systems

- In Course 4 we'll look at some of the state-of-the-art (but reasonably common) approaches that are used to implement recommender systems on the web, e.g. rating prediction:

- And "people who bought X also bought Y" etc.:

Recommender Systems

Other than building such systems in Course 4, in the meantime we're also interested in "standard" tasks that can approached using the same type of data, e.g.

- Regression and classification tasks
- Time series modeling
- Text analysis
- Visualization
- (Etc.)

Summary of concepts

- Introduced the concept of Recommender Systems, which we'll use as an ongoing case-study throughout the Specialization