Machine Learning for Business Processes

Ashish Upadhyay

PhD Student, Computing Science (AI), RGU Aberdeen
Alumnus, BTech (Class of '19), IIIT Naya Raipur

Outline

- Business Processes
- Machine Learning
- Labelled Data
- Active Learning
- Closing Note

Business Processes

- A small process of some big business
- Take an example:
  - Business: Regulatory Management
  - Process: Identify the source of regulatory documents
- Classifying a bunch of documents
  - A Machine Learning task

Image source: https://www.shutterstock.com/search/regulation

Machine Learning

- Learning relation between input and output space
  - Using labelled data
- We need **Data, Labelled Data**!
  - ML algorithms are data hungry
- Getting labelled data
  - Expensive and time consuming
- Biggest question:
  - How do we get labelled data?

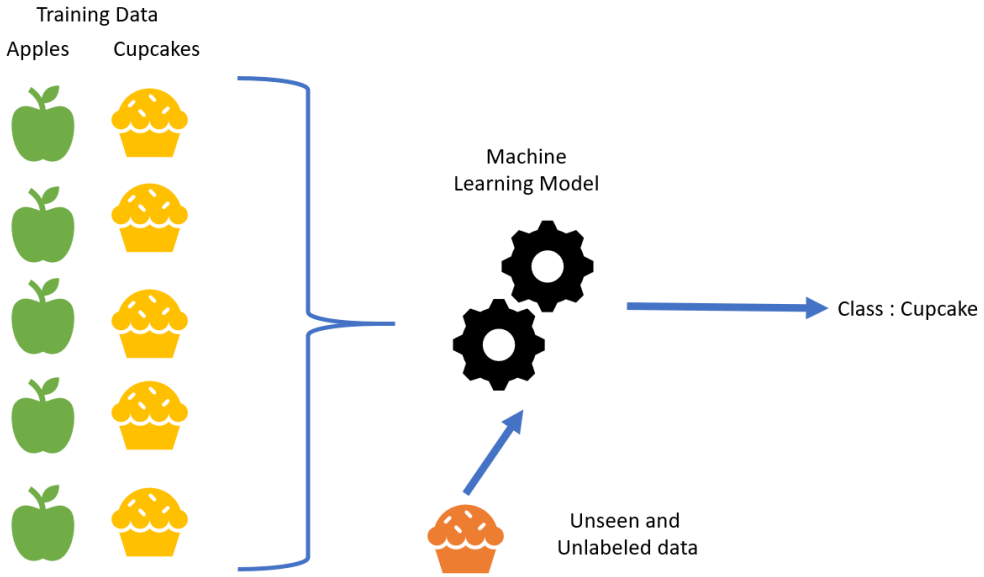


Image source: https://medium.com/@canburaktumer/

Labelled Data

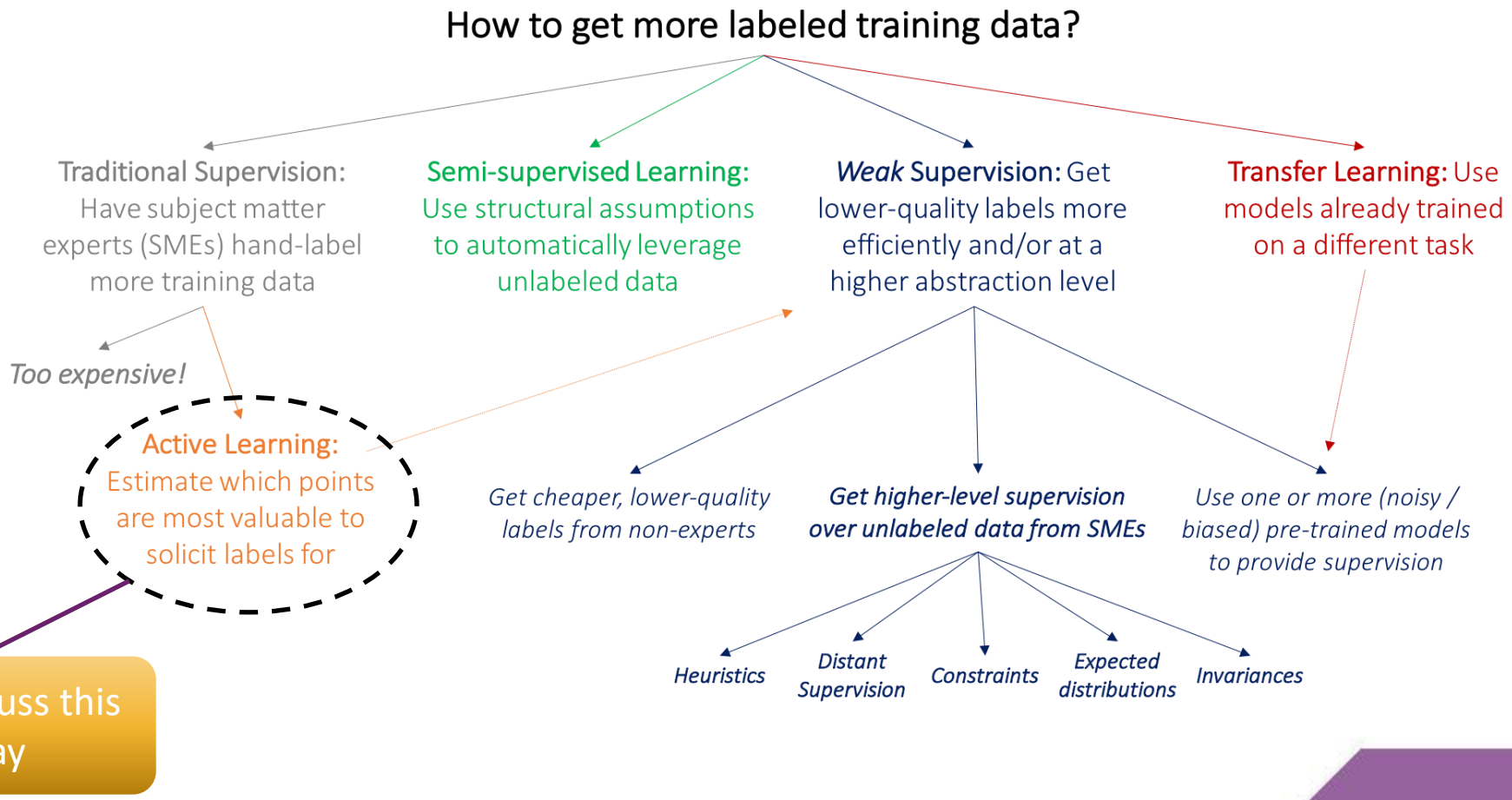


Image source: http://ai.stanford.edu/blog/weak-supervision/

Active Learning: Human-in-the-Loop

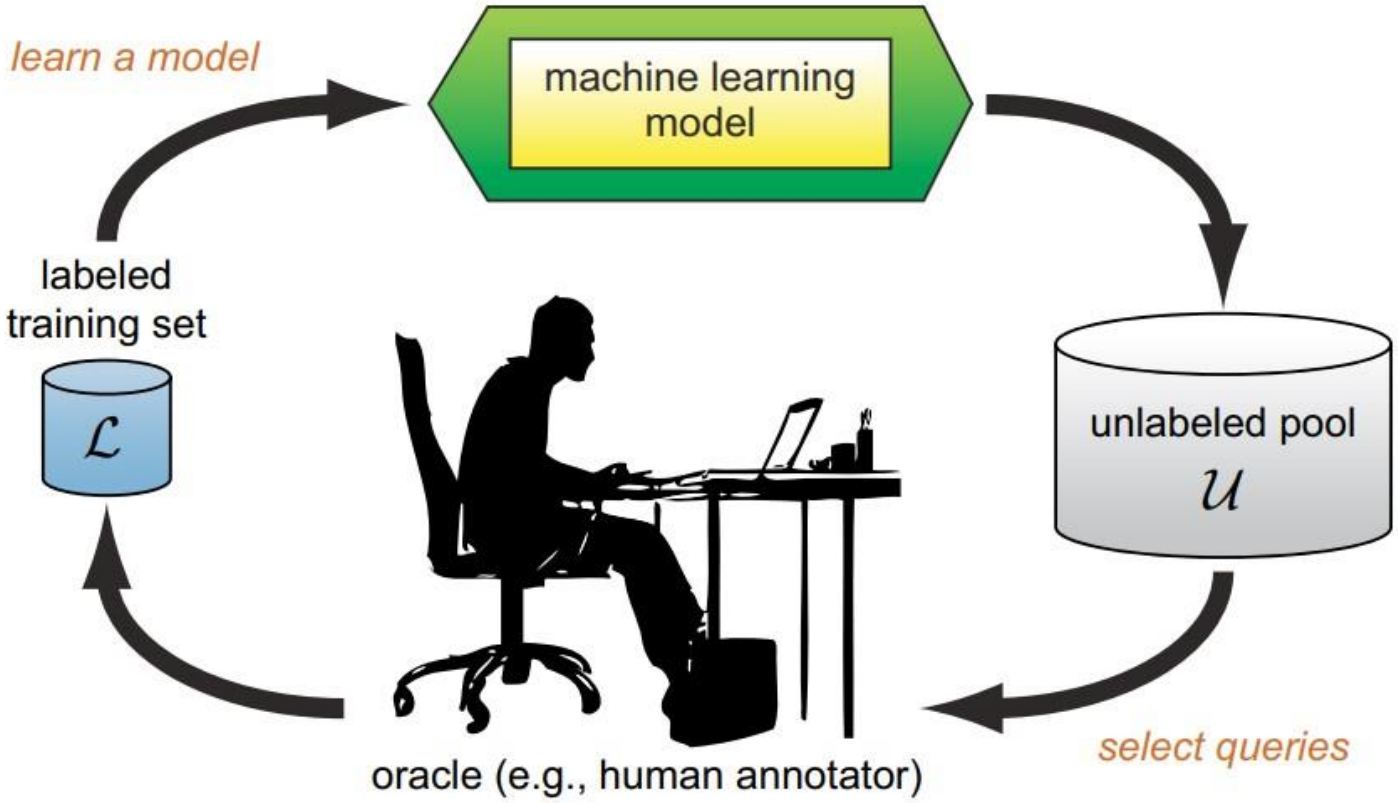
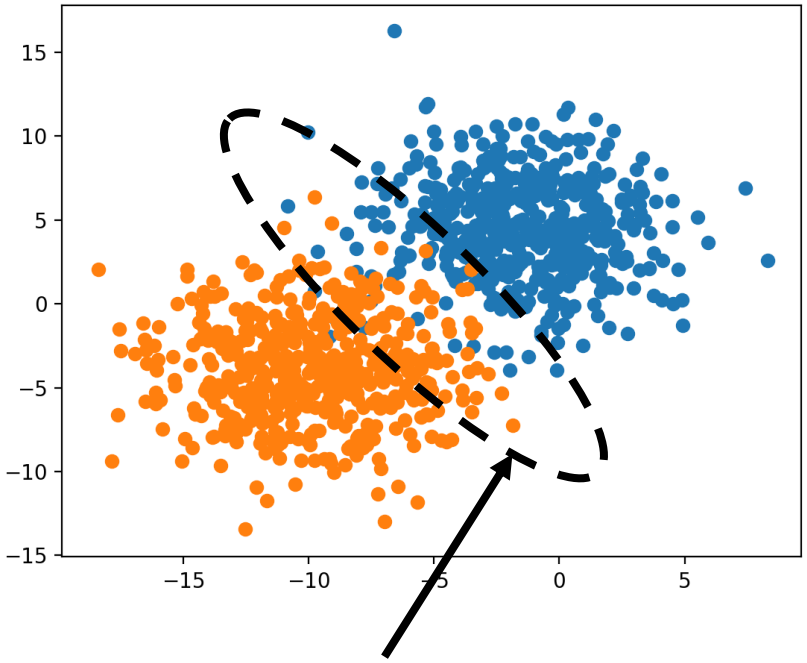


Image source: http://burrsettles.com/pub/settles.activelearning.pdf

Active Learning – An Example

- Assumption
  - We've huge sample of $Unlablled\ data\ (U)$
- Aim
  - Oracle should label I
- **Step 0**: Divide U into
  - $Seed\ (L)$: some labelled samples from U
  - $Unlablled\ pool\ (U)$: rest of the dataset

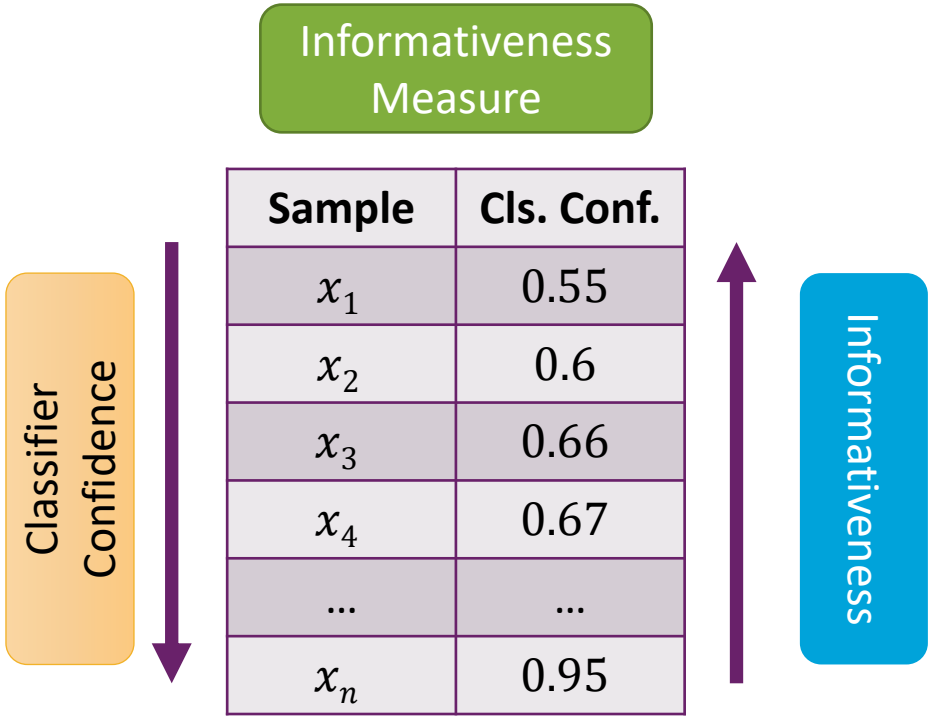


Most informative area (I)

Image source: https://machinelearningmastery.com/

Active Learning – An Example

- **Step 1**: Train a $classifier\ (C)$ on L
  - Classifier based on domain knowledge
- **Step 2**: Use C to rank samples of U
  - Rank based on some $informativeness$
- **Step 3**: Label $informative\ samples$
- **Step 4**: Add new labelled samples to L
- Repeat until stopping criteria



Sample	Cls. Conf.
x_1	0.55
x_2	0.6
x_3	0.66
x_4	0.67
…	…
x_n	0.95

$$x_i \in U, \forall i \in \{1, 2, \ldots, |U|\}$$

Active Learning – Goal

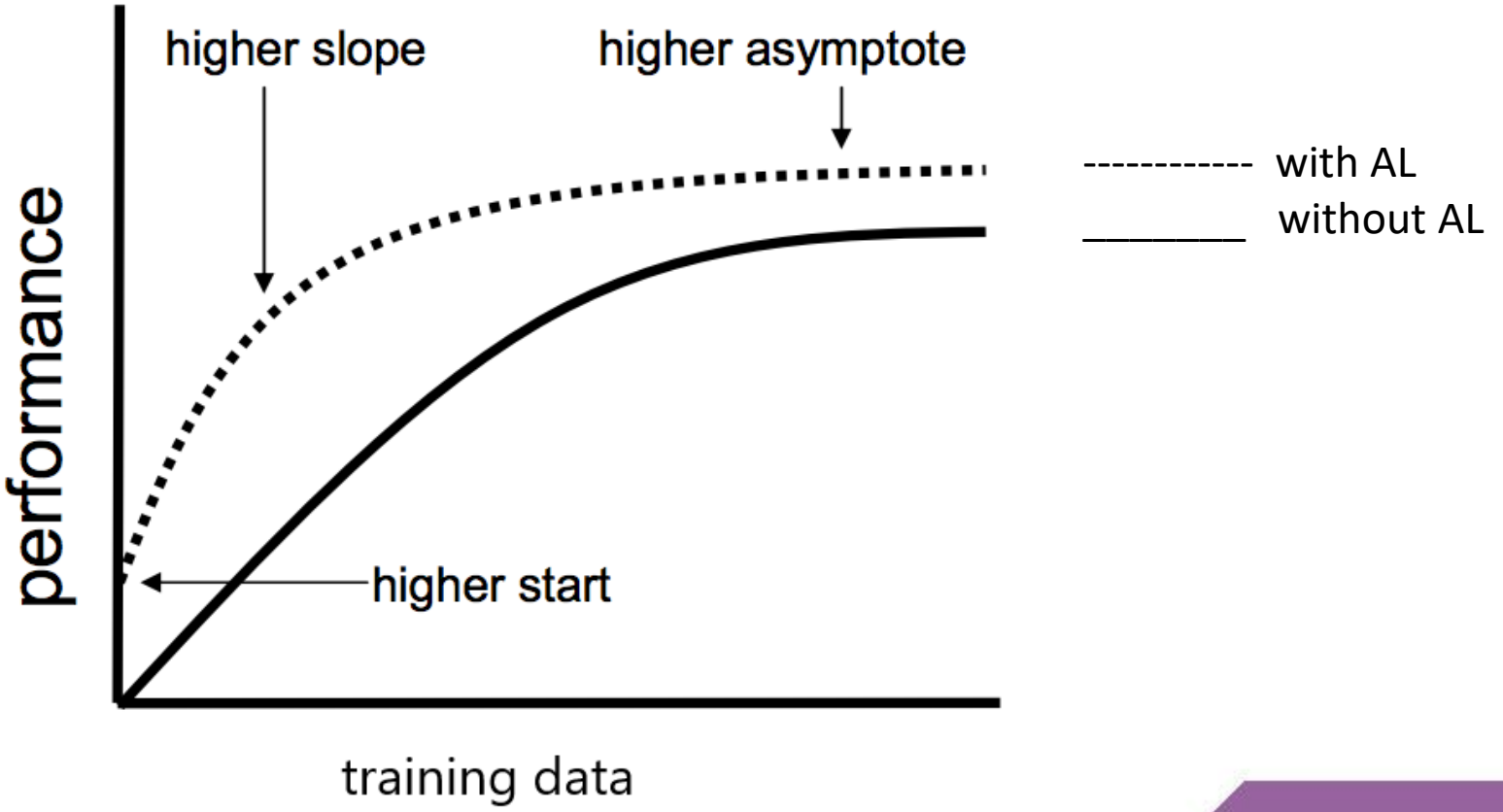


Image source: https://machinelearningmastery.com/

Closing Note

- Thumb rule of Machine Learning
  - Data first, algorithm next
  - Always try Active Learning for manual labelling
- Data preparation on your resume
  - Always set you aside from the crowd
- Must Read:
  - Settles, Burr. **Active learning literature survey**. University of Wisconsin-Madison Department of Computer Sciences, 2009.

Thank You!

- Please feel free to ask any question
- Reach me at:
  - E-mail: ashish.{rgu/iiitnr}@gmail.com
  - LinkedIn: https://www.linkedin.com/in/ashishu007/
- Want to know about my research?
  - I blog at: https://ashishu007.live/blog/
  - Soon, I'll write this talk as well