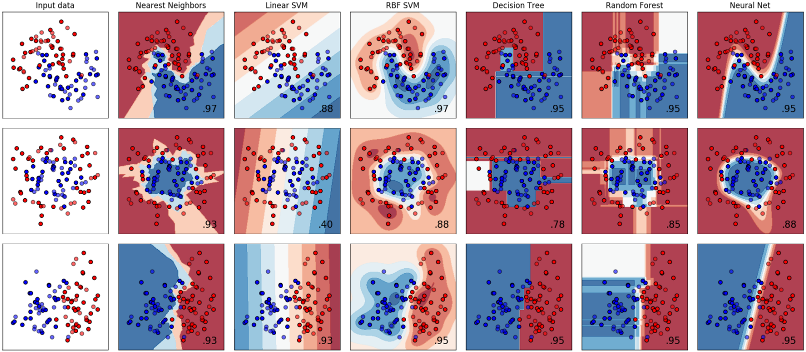
Machine Learning

Introduction

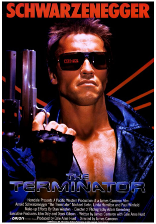
Maxime Gasse

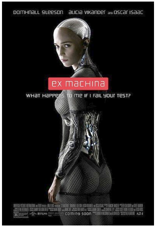


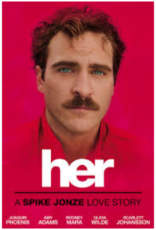
http://scikit-learn.org

Artificial Intelligence

In collective awareness







In scientific research

«~~Can machines think?~~ Can machines do what we can do? »
Alan Turing, 1950

«The building of computer programs which perform tasks which are, for the moment, performed in a more satisfactory way by humans »
Marvin Lee Minsky, 1956

«AI is anything that hasn't been done yet »
Hofstadter, 1980

Machine learning

Good old-fashioned AI (GOFAI, 1950s - 1980s): place knowledge into computers.

Machine learning (1960s - today): let computers acquire knowledge by themselves.

- statistics, mathematical optimization, computer science.

Recent successes

Closely related to deep learning (artificial neural networks).

- Handwritten synthesis: `https://goo.gl/5QEJXn`
- Speech synthesis: `https://goo.gl/RaS6WE`
- Facial animation: `https://youtu.be/lDzrfdpGqw4?t=25s`
- Style transfer: `https://goo.gl/atHKFg`
- Game-playing: AlphaGo, DotA2 `https://youtu.be/wpa5wyutpGc`

Many more...
`https://www.kaggle.com/competitions`

Preliminary materials

Probabilistic framework

Notations

- domain $\mathcal{V}$, e.g. $\mathbb{R} \times \{0, 1\} \times \dots$
- sample $\mathbf{v} \in \mathcal{V}$, e.g. $(v_1, v_2, \dots)$
- data set $\mathcal{D} = \{\mathbf{v}^{(i)}\}_{i=1}^N$

Example (student marks)

$$\begin{aligned}
\mathbf{v}^{(1)} &= (\text{male}, 12.0, 9.5) \\
\mathbf{v}^{(2)} &= (\text{female}, 15.0, 11.0) \\
\mathbf{v}^{(3)} &= (\text{female}, 7.0, 15.0) \\
&\dots
\end{aligned}$$

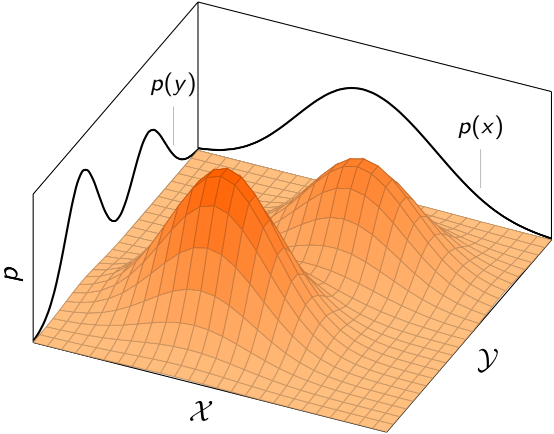
i.i.d hypothesis: independent and identically distributed samples

- independent $p(\mathcal{D}) = \prod_i p(\mathbf{v}^{(i)})$
- identically distributed $\mathbf{v}^{(i)} \sim p_{\mathbf{v}}, \forall i$

Marginal distribution

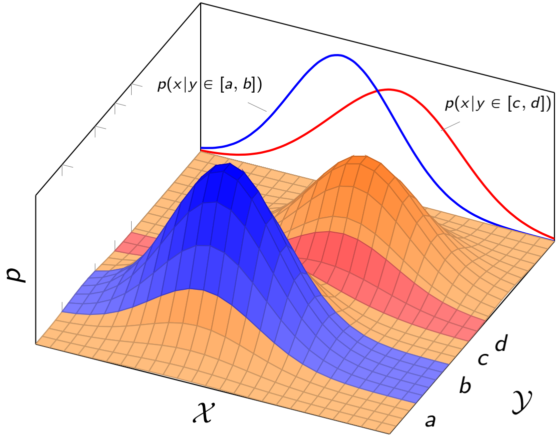
$$p(x) = \sum_y p(x, y)$$

$$\text{(or } \int_y p(x, y)\mathrm{d}y)$$



Conditional distribution

$$p(y|x) = \frac{p(x,y)}{p(x)}$$



Machine learning

Model a probability distribution p from $\mathcal{D}$, then answer queries.

- Supervised learning: $p(\mathbf{y} \mid \mathbf{x})$ for classification, regression, ...
- Unsupervised learning: $p(\mathbf{v})$ for clustering, anomaly detection, sampling, ...
- Reinforcement learning: $p(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$ for game-playing, ...

Many intersections:

$$p(\mathbf{y} \mid \mathbf{x}) = \frac{p(\mathbf{x}, \mathbf{y})}{\sum_{\mathbf{y}'} p(\mathbf{x}, \mathbf{y}')}.$$

Course overview

Lectures: 2 x 2h

- supervised learning (2h30);
- unsupervised learning (1h);
- reinforcement learning (30min).

Practical work: 1 x 4h

- python + scikit-learn;

Evaluation: final exam + practical work.

Lecture sources:

C. M. Bishop (2006). Pattern recognition and machine learning.
I. Goodfellow, Y. Bengio, and A. Courville (2016). Deep Learning.
T. Hastie, R. Tibshirani, and J. H. Friedman (2009). The elements of statistical learning: data mining, inference, and prediction, 2nd Edition.

http://scikit-learn.org