Machine Learning: Multiple Kernel Learning

Gustavo Ramirez

November 18, 2016

Overview

What exactly is Machine Learning?

A lot of things.. and the field is constantly expanding.

What exactly is Machine Learning?

"[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed"

First step towards AI...

Arthur Lee Samuel (1901-1990): computer gaming, artificial intelligence, machine learning, TeX, ...

What exactly is Machine Learning?

"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

Tom Mitchell, Carnegie Mellon University

What exactly is Machine Learning?

- **Supervised machine learning**: The program is trained on a pre-defined set of training examples, which then facilitate its ability to reach an accurate conclusion when given new data.
- **Unsupervised machine learning**: The program is given a bunch of data and must find patterns and relationships therein.

What exactly is Machine Learning?

In supervised ML, two major subcategories are:

- **regression**: fitting
- **classification**: systems where we seek a yes-or-no prediction

What exactly is Machine Learning?

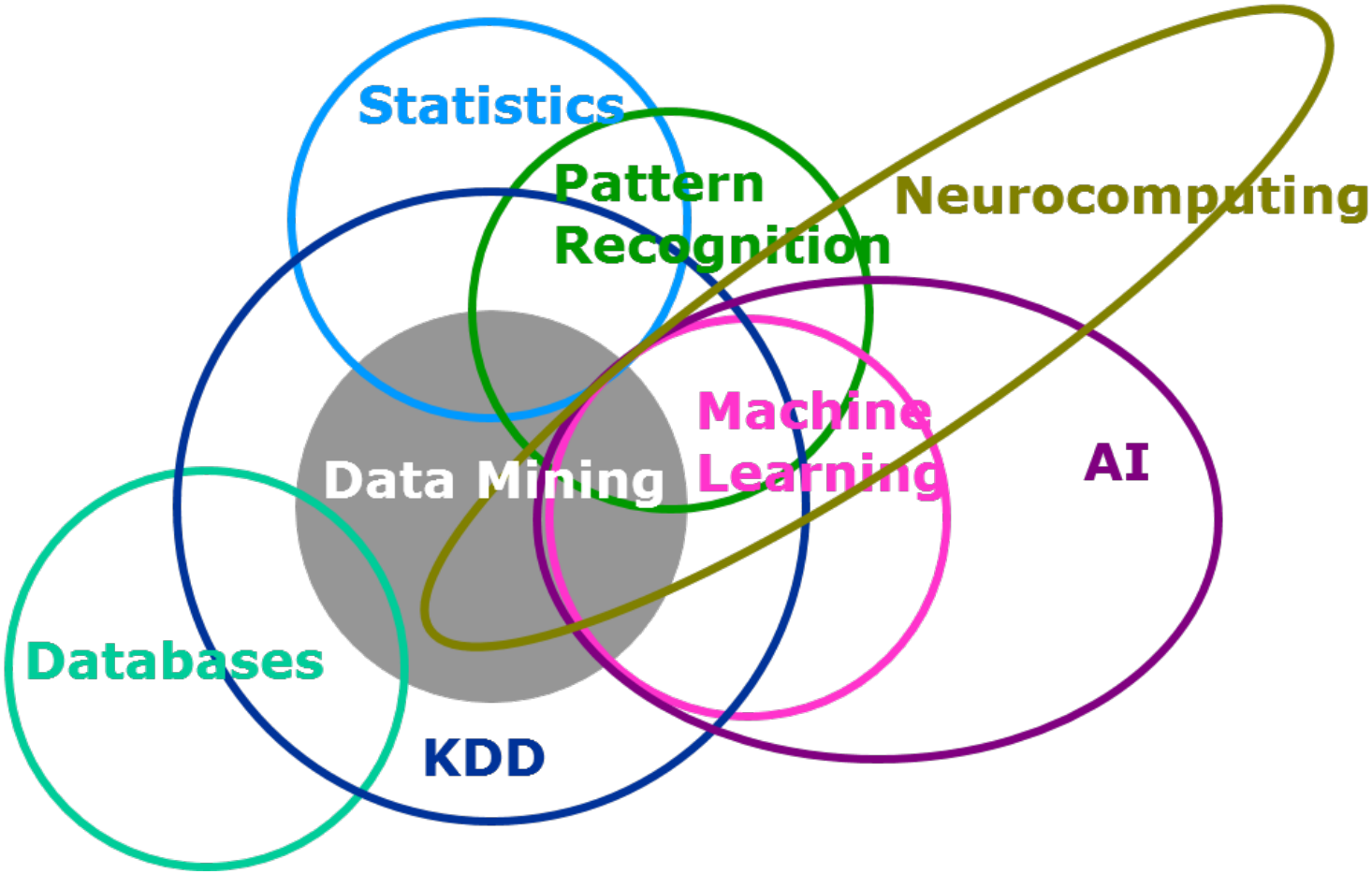


Image taken from: "SAS, Data Mining and Machine Learning".

What exactly is Machine Learning?

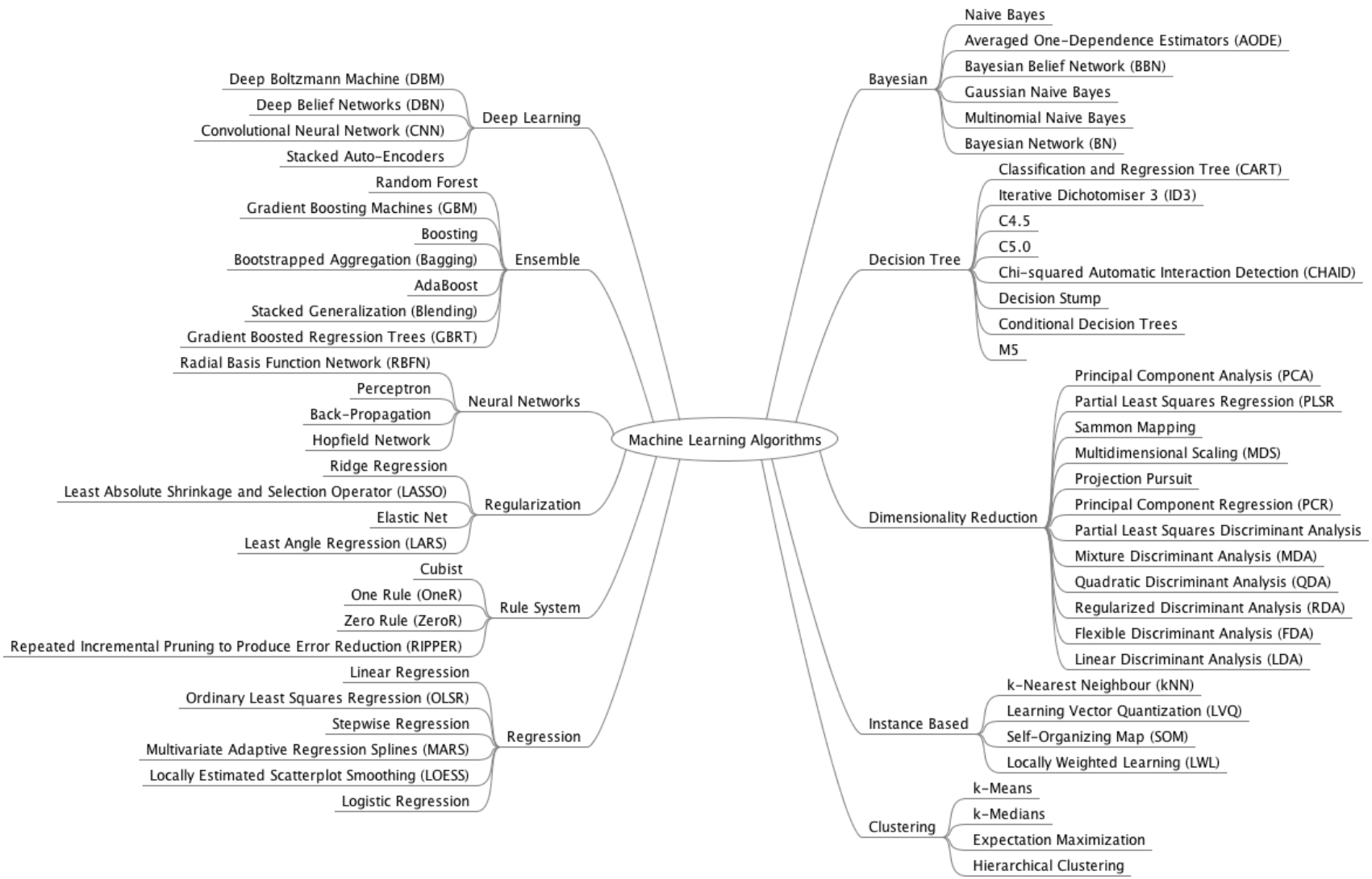


Image taken from: "A Tour of Machine Learning Algorithms".

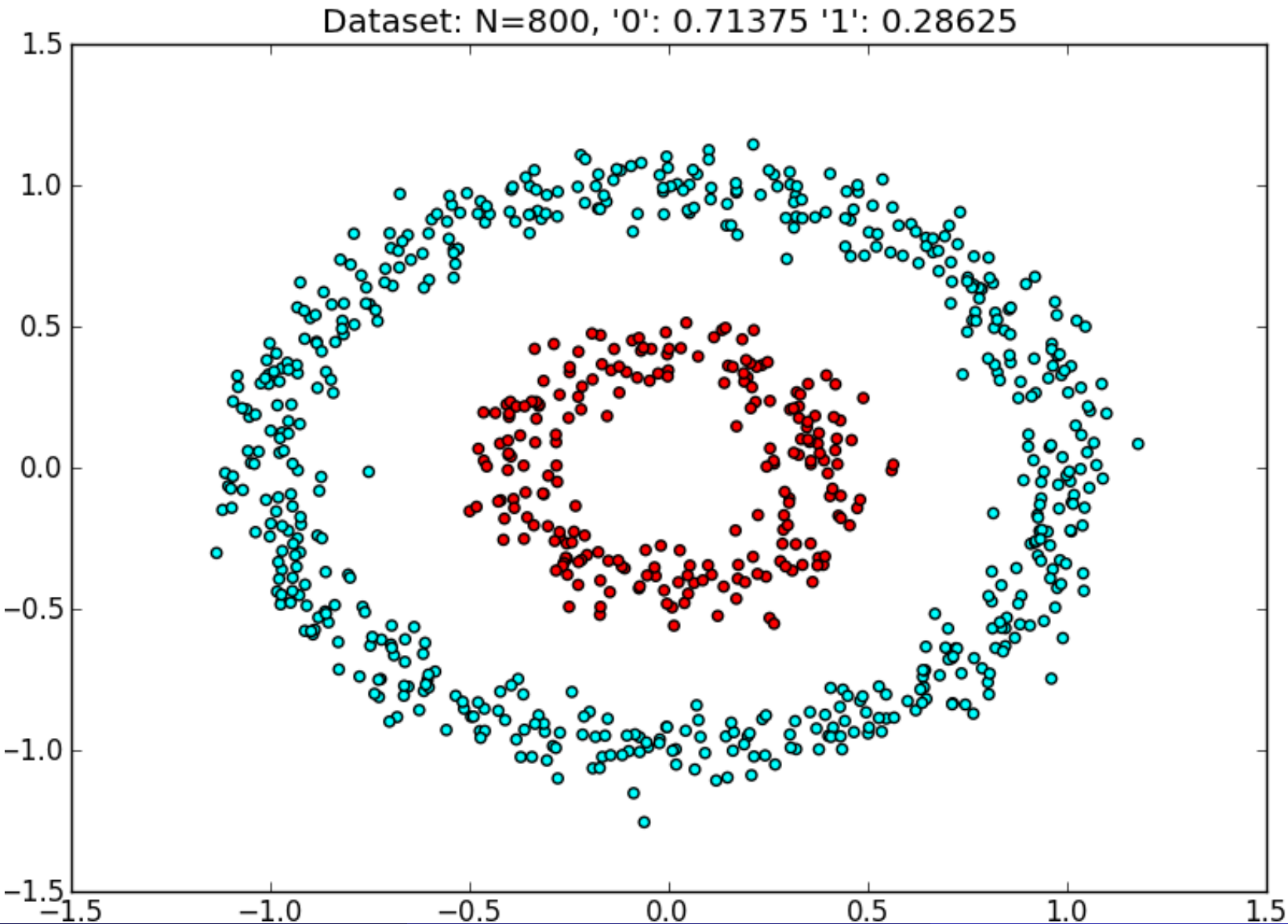
MKL: meaning

MKL:
"set of machine learning methods that use a predefined set of kernels and learn an optimal linear or non-linear combination of kernels as part of the algorithm"

MKL: use in SVM

SVM:
"given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples"

MKL: use in SVM



MKL: use in SVM

The Kernel trick:

- many ML algorithms (e.g. SVM) use the data only through inner products.
- e.g. the following matrix (**kernel matrix**) can be used in classification and regression:

$$X = \begin{bmatrix} \vec{x_1} \\ \vec{x_2} \\ \cdots \\ \vec{x_n} \end{bmatrix} \rightarrow K = XX^T \tag{1}$$

MKL: use in SVM

The Kernel trick:

- the idea is to apply a transformation $\phi(\vec{x})$, which preserves the form of K:

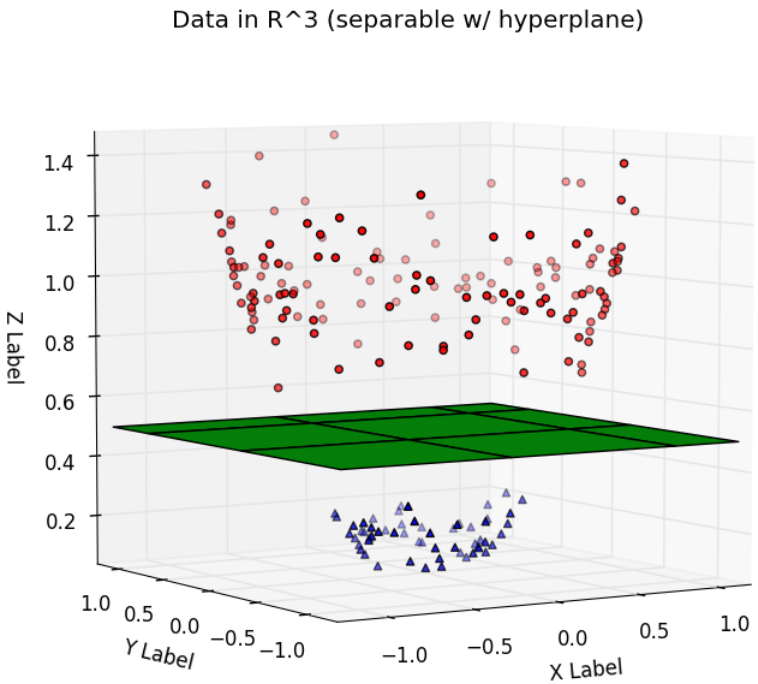
$$K = \begin{bmatrix} \phi(\vec{x_1})^T \phi(\vec{x_1}) & \phi(\vec{x_1})^T \phi(\vec{x_2}) & ... \\ \phi(\vec{x_2})^T \phi(\vec{x_1}) & ... & ... \\ ... & ... & ... \end{bmatrix}$$

MKL: use in SVM

The Kernel trick:

- simple example (on a ϕ preserving the form of K):
  - transformation ϕ: $(x_1, x_2) \rightarrow (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$
  - let's take: $\vec{r} = \phi(\vec{a})$ and $\vec{s} = \phi(\vec{b})$
  - $\Rightarrow (\vec{r} \cdot \vec{s})_{3D} = r_1s_1 + r_2s_2 + r_3s_3 = (a_1^2)(b_1^2) + (\sqrt{2}a_1a_2)(\sqrt{2}b_1b_2) + (a_2^2)(b_2^2)$
  - $\Rightarrow (\vec{r} \cdot \vec{s})_{3D} = (\vec{a} \cdot \vec{b})^2$
- then, the Kernel trick consists of only knowing how to compute the inner product in the new space (through a function of the inner product in the original space), but not the actual transformation

MKL: the Kernel trick



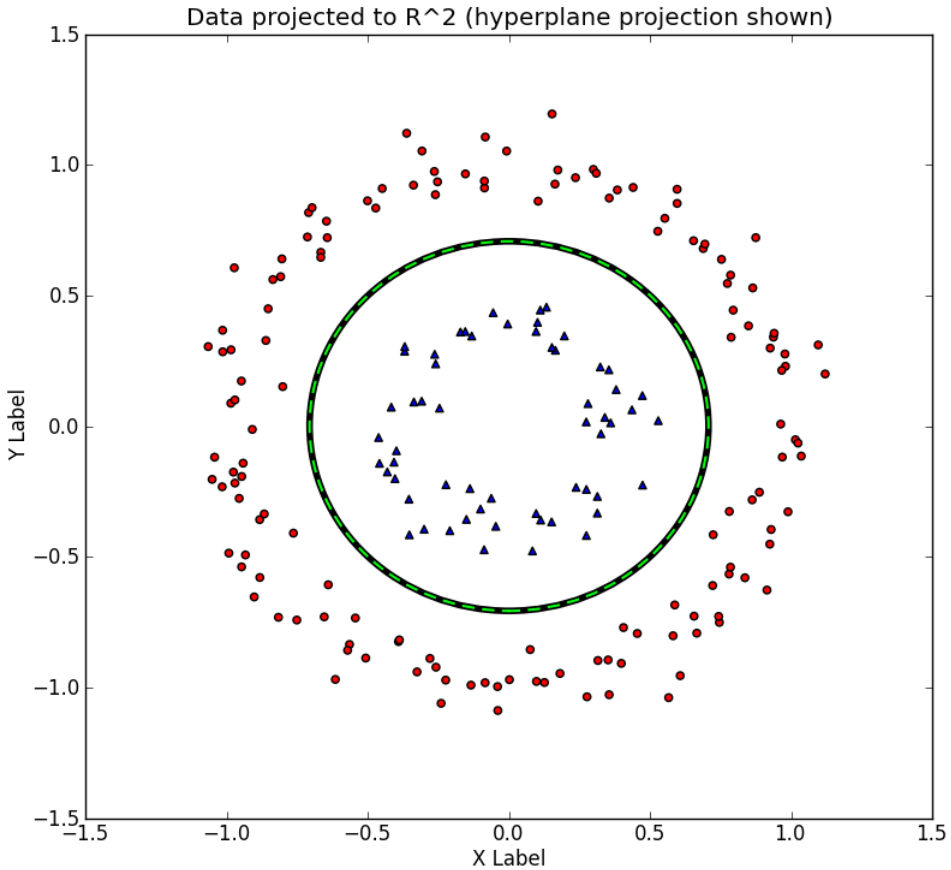


Image taken from: "The Kernel Trick".

MKL: the Kernel trick

MKL:

- extension of the Kernel trick, using many ϕ's now, and optimize with the original algorithm (SVM or other) but taking those kernels into account

Applications of ML

Prediction in cancer:

Published in final edited form as:

Nat Biotechnol. 2014 December ; 32(12): 1202–1212. doi:10.1038/nbt.2877.

A community effort to assess and improve drug sensitivity prediction algorithms

James C Costello[1,2,13,14], **Laura M Heiser**[3,14], **Elisabeth Georgii**[4,14], **Mehmet Gönen**[4], **Michael P Menden**[5], **Nicholas J Wang**[3], **Mukesh Bansal**[6], **Muhammad Ammad-ud-din**[4], **Petteri Hintsanen**[7], **Suleiman A Khan**[4], **John-Patrick Mpindi**[7], **Olli Kallioniemi**[7], **Antti Honkela**[8], **Tero Aittokallio**[7], **Krister Wennerberg**[7], **NCI DREAM Community**, **James J Collins**[1,2,10], **Dan Gallahan**[11], **Dinah Singer**[11], **Julio Saez-Rodriguez**[5], **Samuel Kaski**[4,8], **Joe W Gray**[3], and **Gustavo Stolovitzky**[12]

[1]Howard Hughes Medical Institute, Boston University, Boston, Massachusetts, USA [2]Department of Biomedical Engineering, Boston University, Boston, Massachusetts, USA [3]Department of Biomedical Engineering, Oregon Health and Science University, Portland, Oregon, USA [4]Helsinki

Applications of ML

Recommender systems (Netflix, Amazon, etc.), data mining, image recognition, etc.

References

James C Costello, *et al.* (2014)

A community effort to assess and improve drug sensitivity prediction algorithms

Nature Biotechnology 32, 12021212

Sören Sonnenburg, *et al.* (2006)

Large Scale Multiple Kernel Learning

Journal of Machine Learning Research 1531 – 1565.

The End