

Machine Learning Basics

Shusen Wang

Learning from Data



Test Data



Train

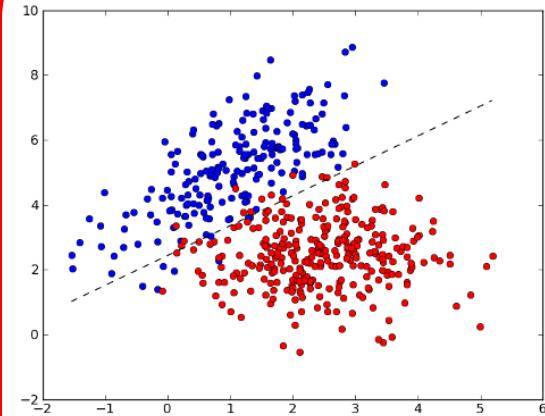


Predict
Data (and perhaps labels)

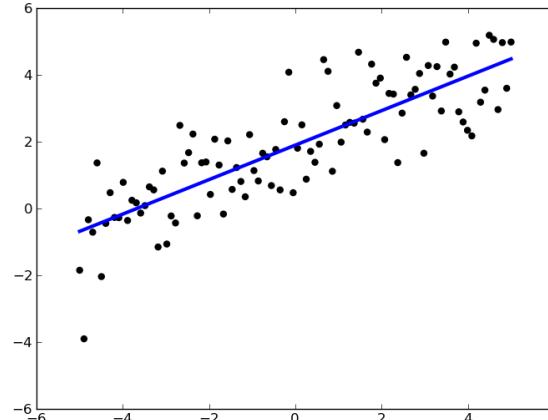
A vertical stack of text and images. To the left of the "Predict" label is a long upward-pointing arrow. To the right of the "Data" label is a blue-outlined box containing the word "car".

Machine Learning Tasks

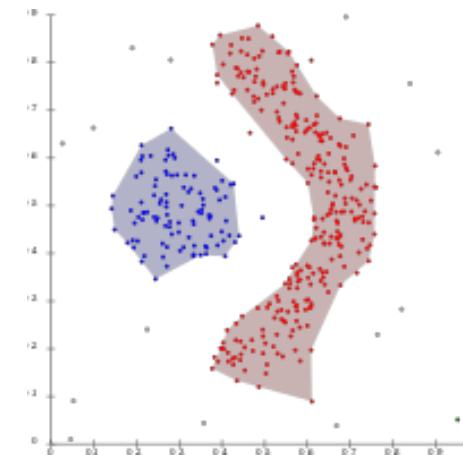
Broad Machine Learning Tasks



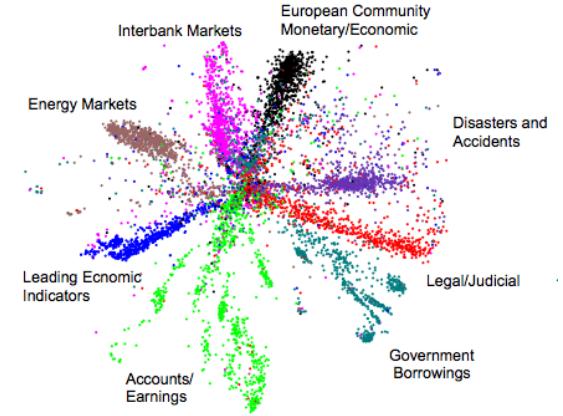
Classification



Regression



Clustering



Dimensionality
Reduction

spam
detection

face
recognition

...

Classification: Spam Detection

The screenshot shows a web-based email interface for 'bConnected powered by Google'. The search bar at the top contains the query 'in:spam'. A yellow banner above the message list encourages enabling desktop notifications for UC Berkeley Mail, with links to 'Learn more' and 'Hide'.

The left sidebar lists various mail categories: Mail (selected), COMPOSE, Inbox, Starred, Sent Mail, Drafts, All Mail, Less ▲, Important, Chats, and Spam (78). Below the sidebar, there is a user profile for 'Shusen' with a blue profile icon and a green dot indicating online status, along with a '+' button.

The main area displays a list of 78 spam messages. Each message is shown in a row with a checkbox, a star icon, the sender's name, and a preview of the message content. The messages include:

- 贡蝎: 业务员为什么有激情, 没有业绩? - shusen 为什么业务员经常申请政策, 但就是没有业绩? 为
- Amazon.com: Dr. Brown's Designed To... - We have a recommendation for you Amazon.com Your Amazon
- Western Dental & Orthodo.: 🎃 We Are Open Today, Celebrate SmileFest - Logo 866-469-5955 Come celebrate with us c
- LinkedIn Salary: Shusen, see salaries for professionals like you. - LinkedIn Shusen Wang Compare salaries
- 富珑: 业务员为什么有激情, 没有业绩? - shusen 为什么业务员经常申请政策, 但就是没有业绩? 为
- Western Dental & Orthodo.: Happy Halloween, We Are Open October 31 🎃 - Logo 866-469-5955 Come celebrate with u
- kqaqy2: Conference to-shusen - P80806PN840XJRF20V8HLFP0V8V6206088J8P8BX0JX4NFDB4(
- IMCIC 2018: 2nd CFP - Complexity, Informatics and Cybernetics - Dear Shusen Wang, Please consider
- Amazon.com: Amazon customers get 2 FREE audiobooks today. - Get a Prime member exclusive offer: 4
- ICPSA: Call for Manuscripts in the field of Probability and Stochastic Analysis - The 3rd Int'l Con
- LinkedIn Editors: Why work stress isn't all bad, and other career trends this week - New: Career tips, news,

Classification: Spam Recognition

Sir / Madam,

We invite you to submit your manuscript(s) for publication. The journals include research papers, review articles, technical projects and short communications containing new insight into any aspect of the covered scope of the journal. Our objective is to inform authors of the decision on their manuscript(s) within weeks of submission. After acceptance, the paper will be published in the current issue immediately.

Keywords: English, Literature, Science, Economics, Engineering, Management, Agriculture, Horticulture, Environment

[International Journal of Advanced Engineering Research and Science \(IJAERS\)](#) ISSN: 2456-1908(O) | 2349-6495 (P)

DOI (CrossRef): [10.22161/ijaers](https://doi.org/10.22161/ijaers)

Thomson Reuters ResearcherID: P-3738-2015

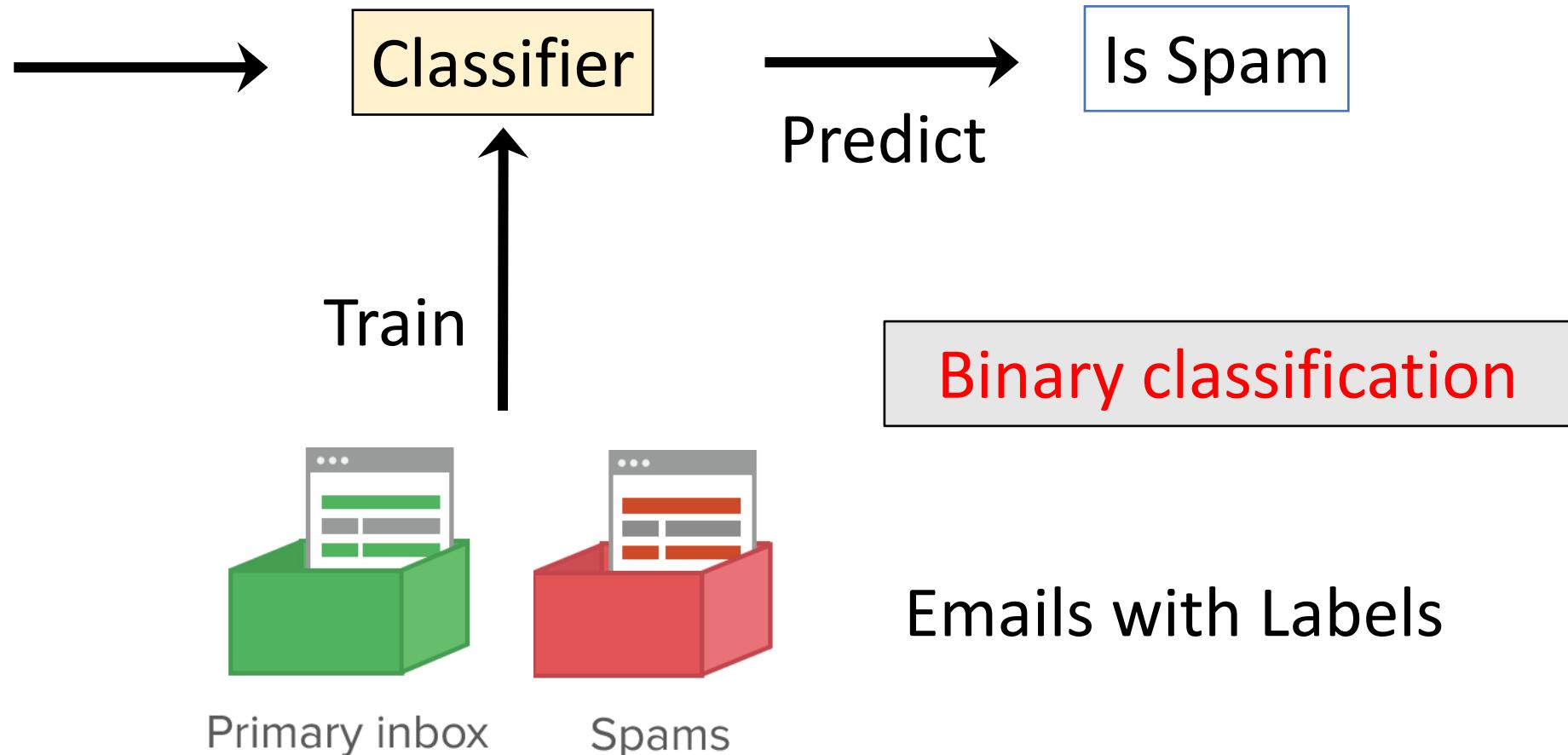
Impact Factor: 4.192, SJIF: 4.072, IBI: 3.2, PIF: 2.465, ISRA-JIF: 1.317,

Website: <http://www.ijaers.com>

Kindly submit research articles to <http://ijaers.com/submit-paper/> or mail us at editor.ijaers@gmail.com

[International Journal of English, Literature and Science \(IJELS\)](#)
ISSN: 2456-7620

New Email



Classification: Spam Recognition

✉ Primary ⚙ Social 🎉 Promotions

✉ **Inbox** 1,195

★ Starred
⌚ Snoozed
➤ Sent
📄 Drafts
✉ All Mail
^K Less
➤ Important
💬 Chats
❗ **Spam** 141

✉ **Shutterfly** [Ad] ENJOY UP TO 50% OFF - Woven Photo Blankets: Upload Your Own Design W

✉ **Canvas Champ** [Ad] Best Online Photo Printing Service - Best Online Photo Printing Service Trus

✉ **Walmart Fashion** Fill their holidays with festive kids' fashion - Find last-minute gifts, party

✉ **eBay** Spotted: Falling prices that have got us excited! - Wow, so much newness! \$

✉ **Costco Wholesale** Dinner is Served! Same-Day Delivery on all Your Holiday Dinner Essentials.

✉ **IKEA USA** Friend, in need of a last minute gift? - Give an IKEA Gift Card View in Brow

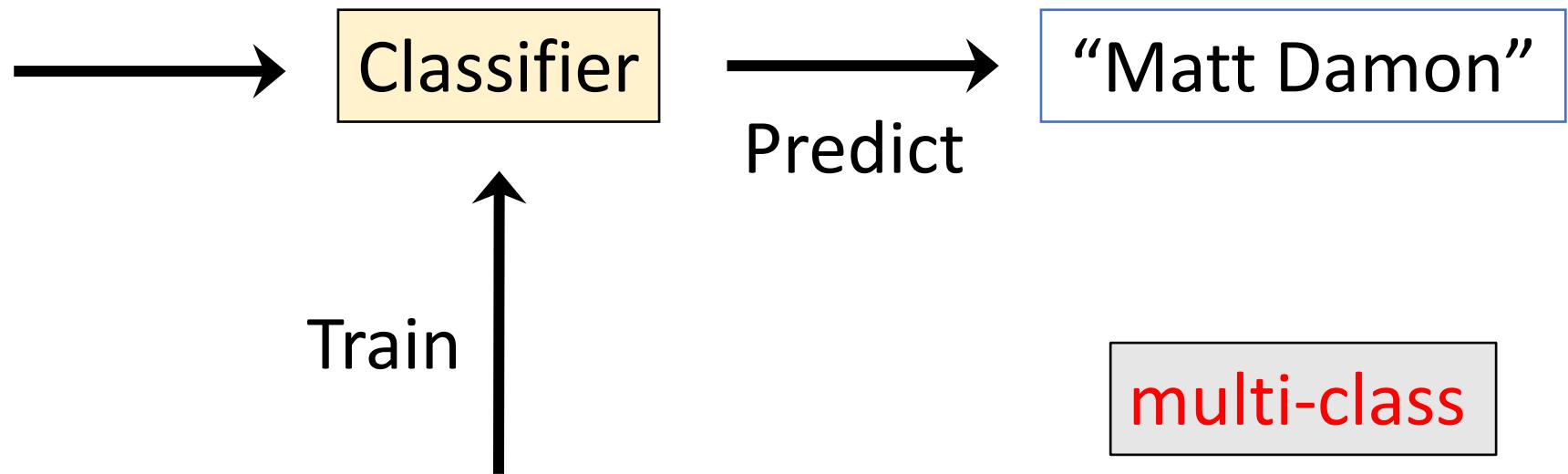
✉ **Walmart** LAST CHANCE to save up to 56% - Here's your final chance to get 1

✉ **Costco Wholesale** LAST HOURS to order in time for Christmas delivery. - Gift ideas - Mac, PCs

✉ **Walmart** HURRY! Last call for free 2-day shipping - Order by Dec. 20 at 2pm for deliv

✉ **Hilton Honors** 'Tis the season for unlimited Points - register now - Start planning your 201

Classification: Face Recognition



Faces & Names

multi-class

Classification: Image Recognition



Test Image



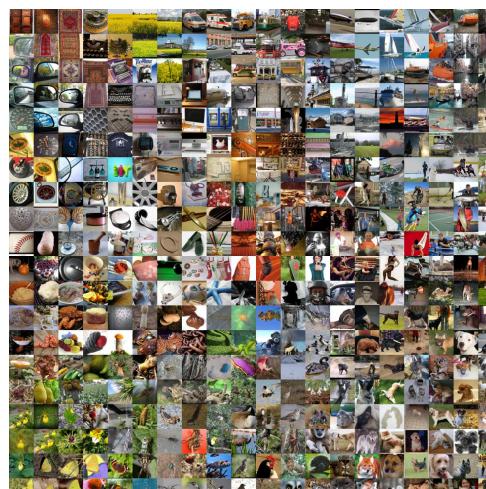
Classifier

Train



Predict

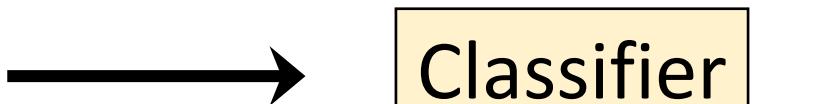
“Dog”
“Puppy”
“Shiba Inu”



multi-class, multi-label

Images & Labels

Classification: Image Recognition



Classifier



Predict

Train

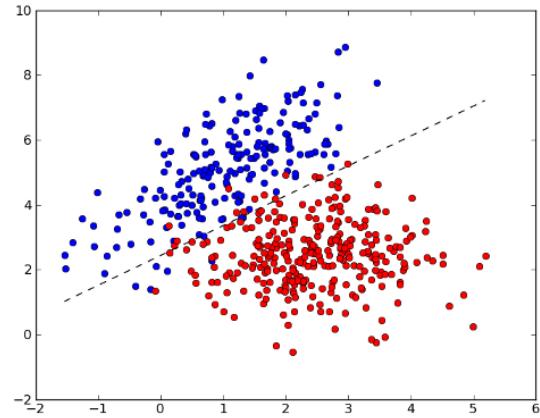


“Dog”
“Shiba Inu”
“Husky”
“Human”
“Woman”

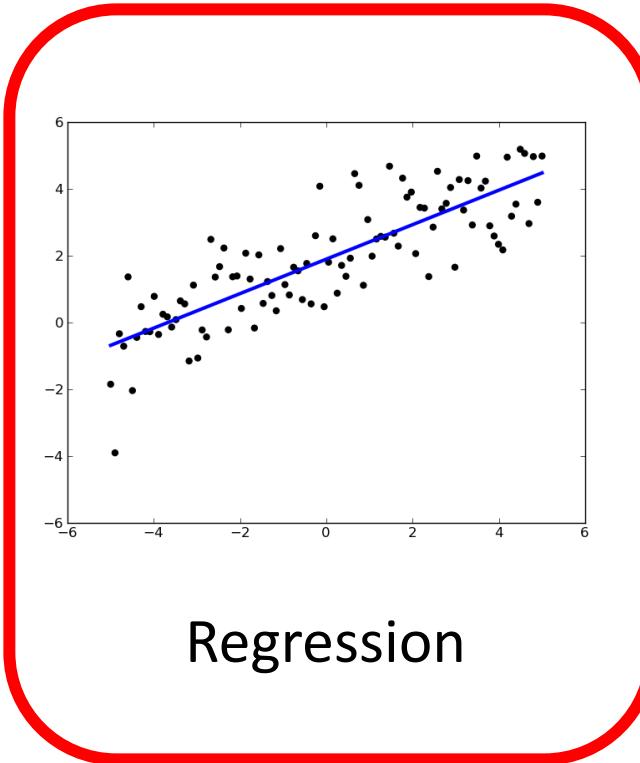
multi-class, multi-label

Images & Labels

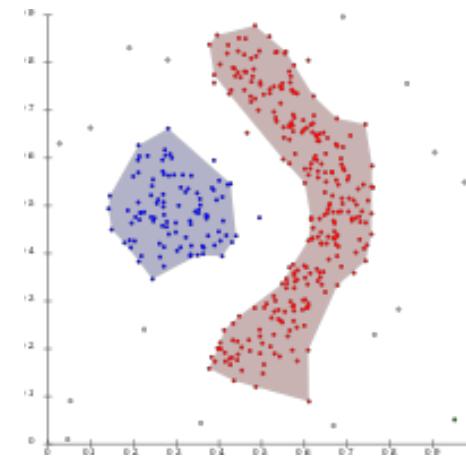
Broad Machine Learning Tasks



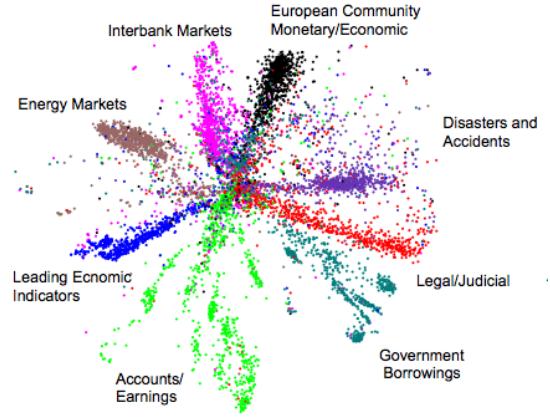
Classification



Regression



Clustering



Dimensionality
Reduction

Regression: Housing Price



Train

label

features

- Price
- Beds
- Baths
- Square Feet
- Miles to Resort
- Miles to Base
- Acres
- Cars
- Years Old
- DoM

Selling price of the property (\$1,000)
Number of bedrooms in the house
Number of bathrooms in the house
Size of the house in square feet
Miles from the property to the downtown resort area
Miles from the property to the base of the ski resort's mountain
Lot size in number of acres
Number of cars that will fit into the garage
Age of the house, in years, at the time it was listed
Number of days the house was on the market before it sold

Regression VS Classification

- **Regression: labels are continuous and ordered**
- Example:

House Prices:

\$324K < \$521K < \$1.2M



Regression VS Classification

- Regression: labels are continuous and ordered
- **Classification: labels are categorical**
- Example:

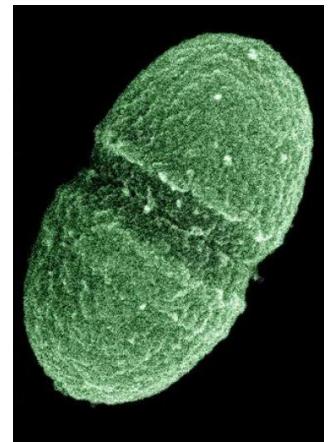
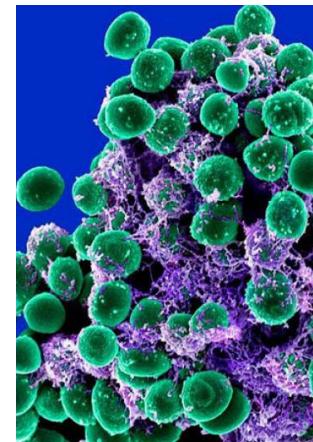
Categories:

Class 1

Class 2

Class 3

Class 4



Regression VS Classification

- Regression: labels are continuous and ordered
- **Classification: labels are categorical**
- Example:

Categories:

Class 1

Class 2

Class 3

Class 4

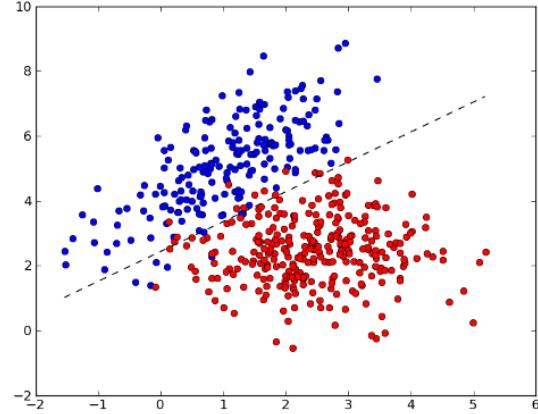
“Class 1” and “Class 4” are categories. They are not ordered!

Class 1

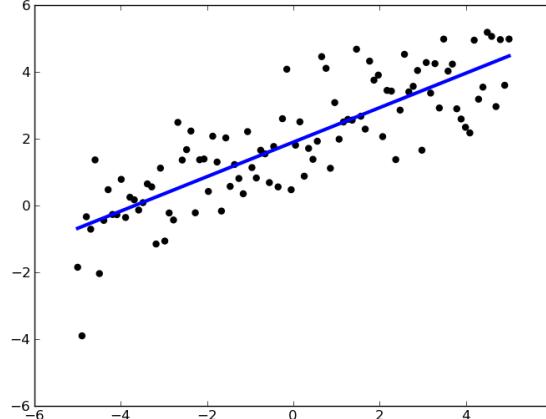


Class 4

Broad Machine Learning Tasks

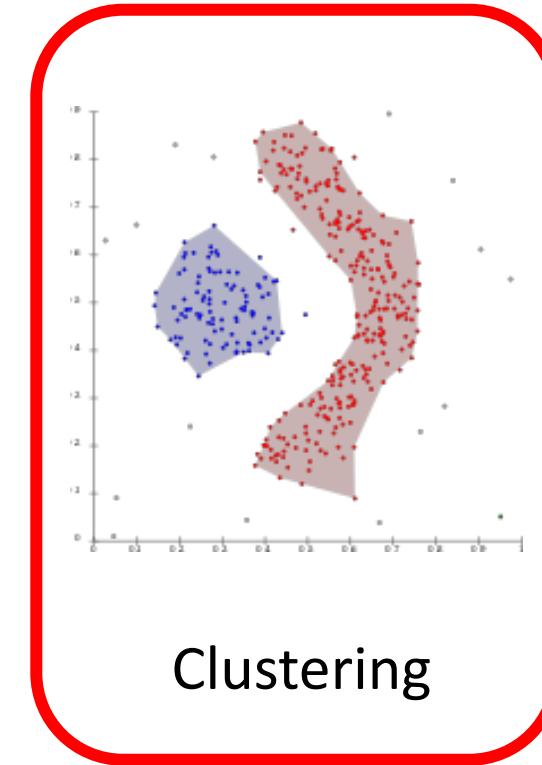


Classification



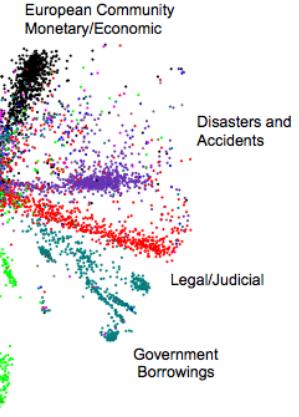
Regression

Supervised Learning



Clustering

Unsupervised Learning



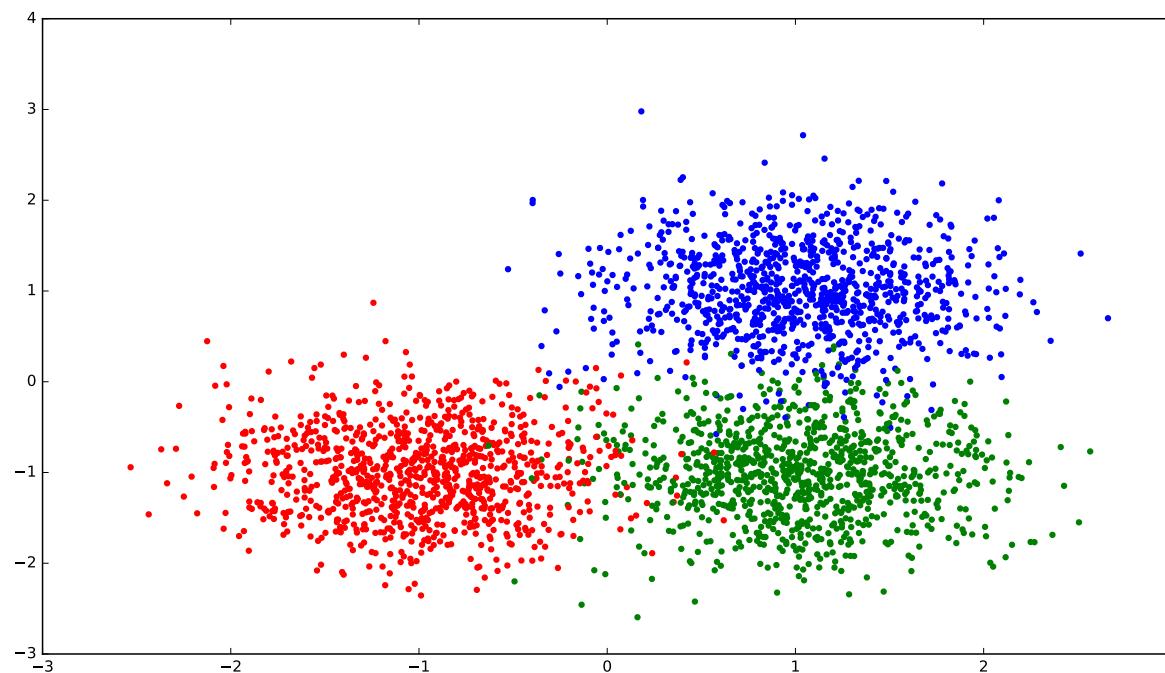
Dimensionality
Reduction

Unsupervised Learning
Supervised Learning

Clustering: Unsupervised Learning

Input: vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ and number of clusters k ($\ll n$).

Output: labels $y_1, \dots, y_n \in \{1, 2, \dots, k\}$.



Clustering: Unsupervised Learning

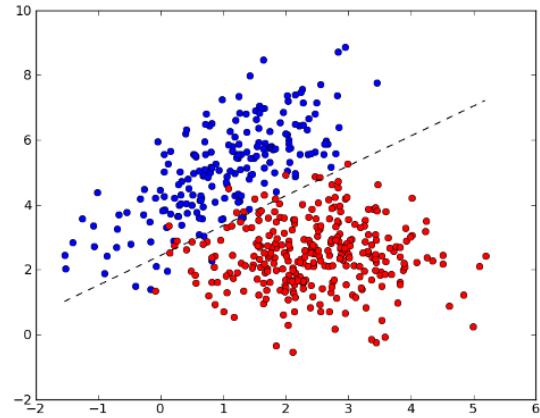
Supervised learning : learn function from **labeled** training data.

- Classification and regression are supervised learning tasks.

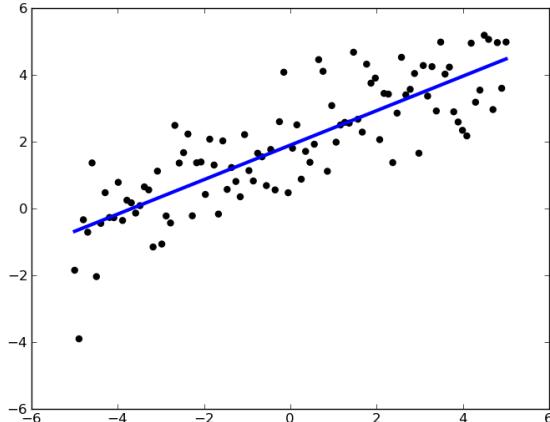
Unsupervised learning : draw inferences from datasets consisting of data **without label**.

- Clustering is an unsupervised learning task.

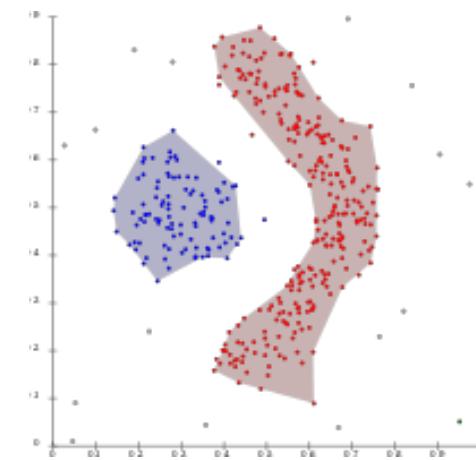
Broad Machine Learning Tasks



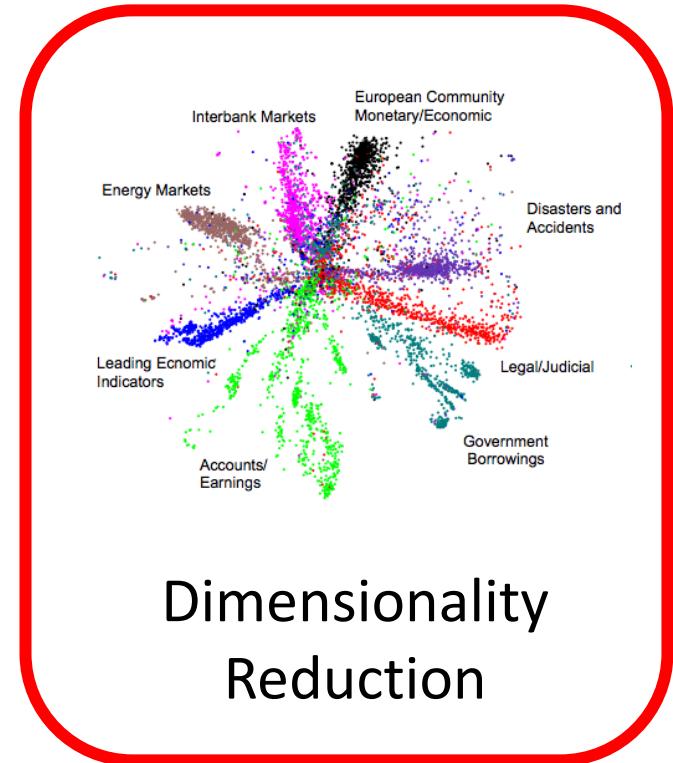
Classification



Regression

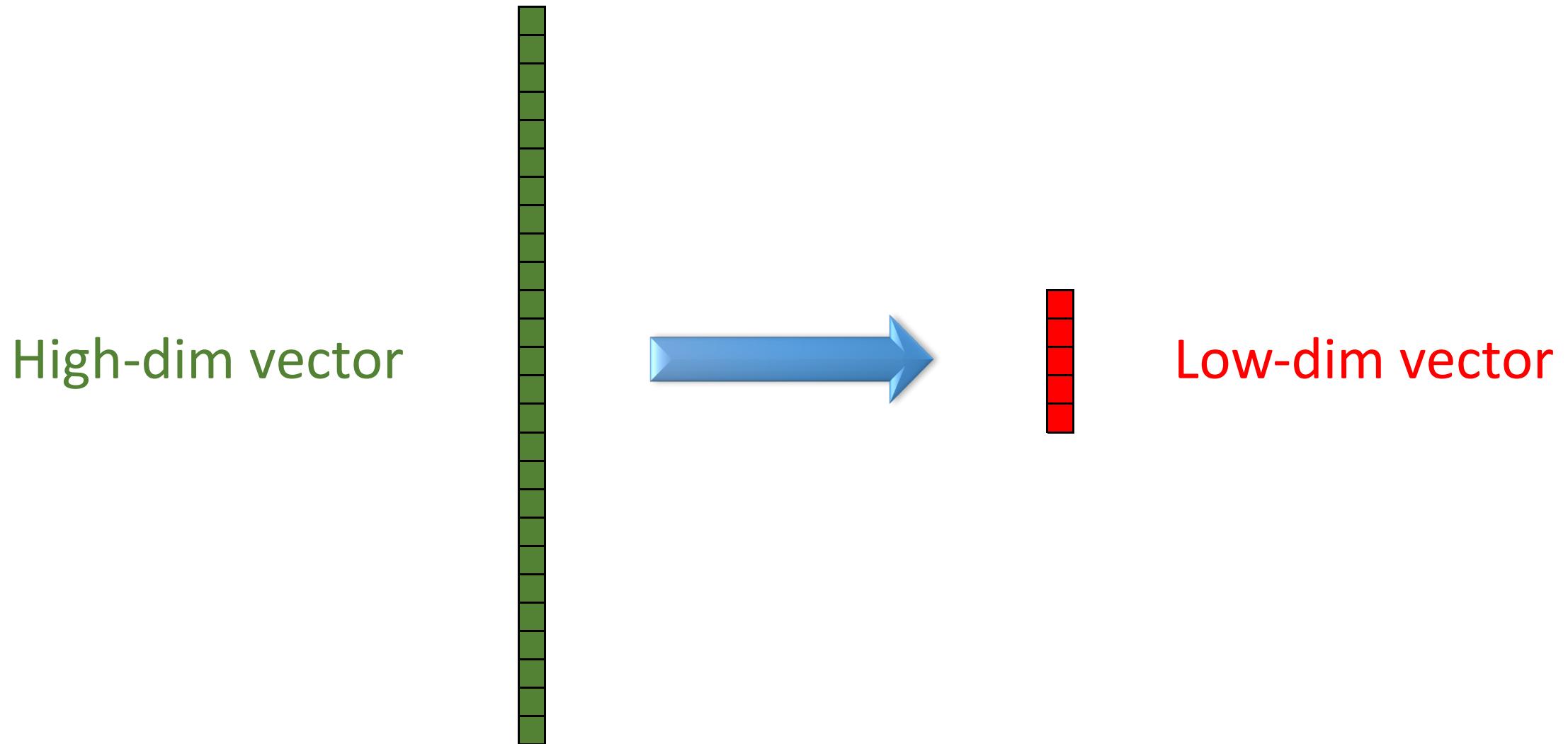


Clustering



Dimensionality
Reduction

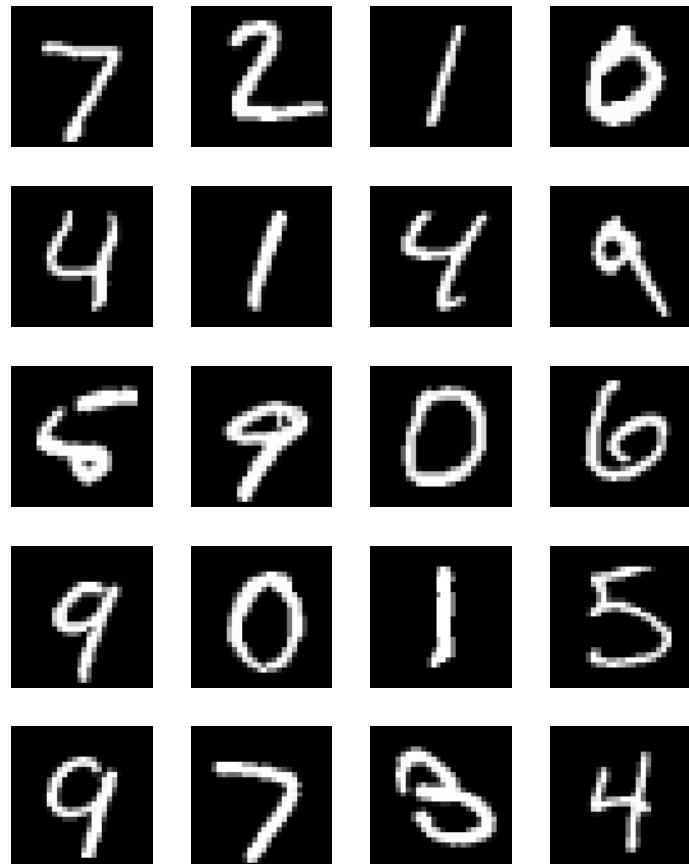
Dimensionality Reduction



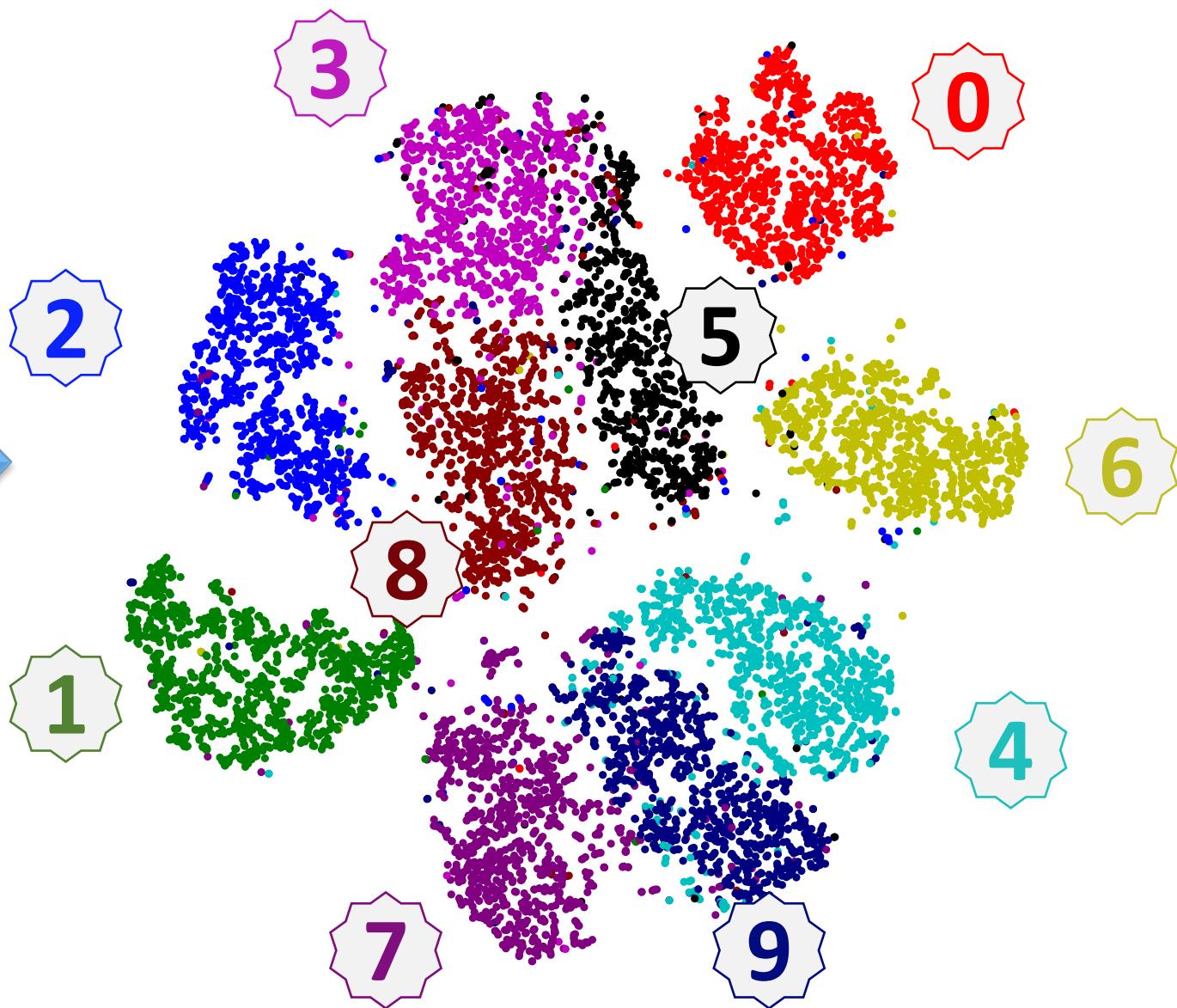
Dimensionality Reduction

- Applications
 - Visualization and analysis
 - Data processing (to make downstream ML more efficient or accurate)
- Methods
 - Principal component analysis (PCA)
 - Kernel PCA
 - Manifold learning
 - Autoencoder
 - Linear discriminant analysis (LDA)

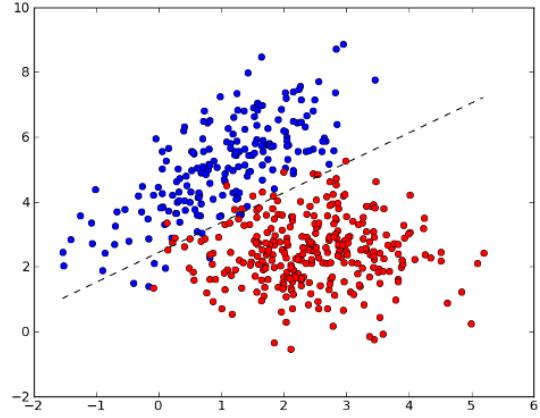
Autoencoder for Hand-Written Digit Data



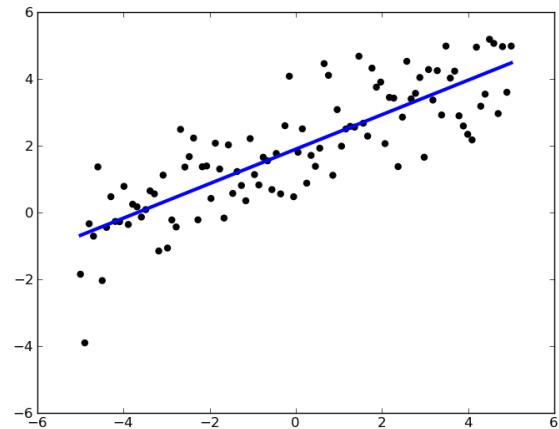
60K images (size 28×28)



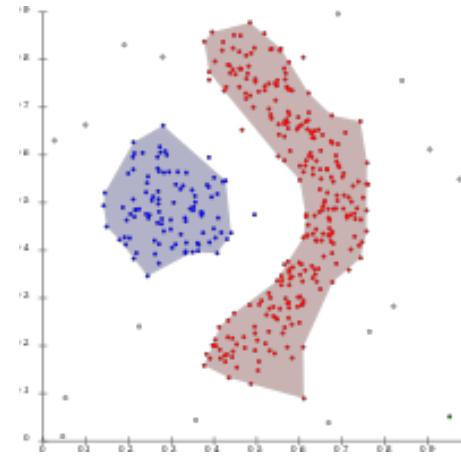
Broad Machine Learning Tasks



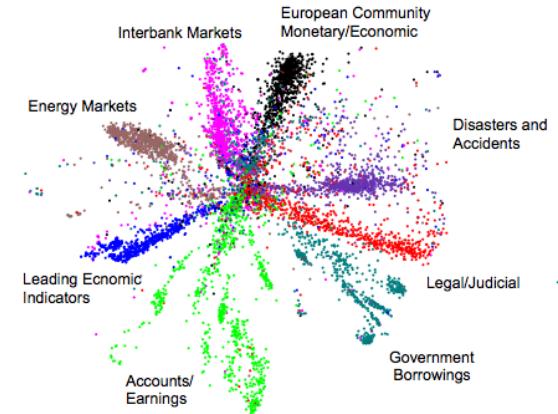
Classification



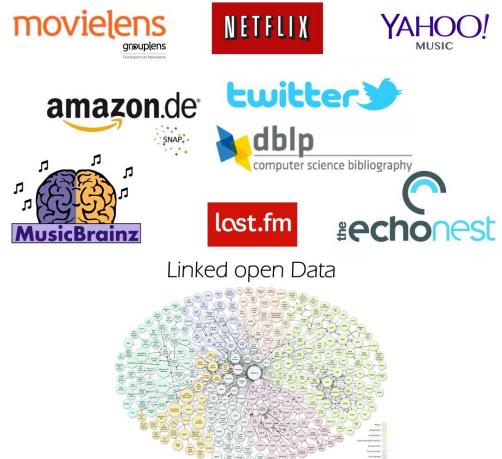
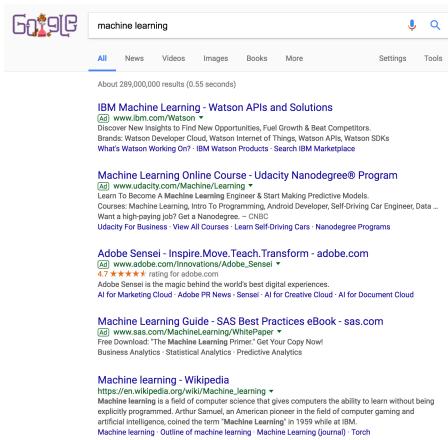
Regression



Clustering



Dimensionality Reduction



Recommendation



Reinforcement Learning

Ranking

Tasks, Methods, & Algorithms

Tasks (or problems)

Regression

Classification

Clustering

Dim Reduction

Tasks (or problems)

Regression

Classification

Clustering

Dim Reduction

Methods (or models)

Neural Networks

SVM

Logistic Regression

Decision Tree

Nearest Neighbor

Tasks (or problems)

Regression

Classification

Clustering

Dim Reduction

Methods (or models)

Neural Networks

SVM

Logistic Regression

Decision Tree

Nearest Neighbor

Algorithms

Gradient Descent (GD)

Stochastic GD

Coordinate Descent (CD)

Dual CD

Newton's Method