

Machine Learning Basics

Shusen Wang

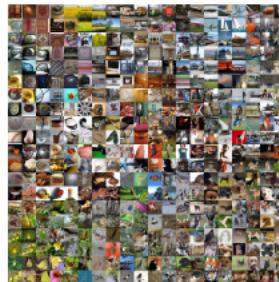
Learning from Data



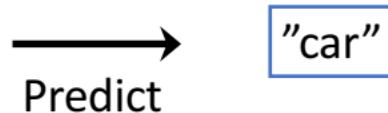
Test Data



Train

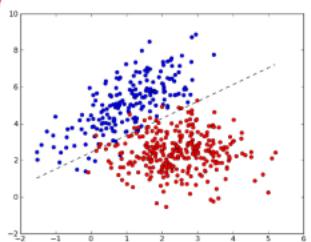


Data (and perhaps labels)

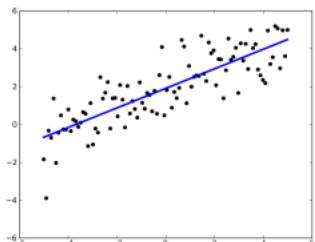


Machine Learning Tasks

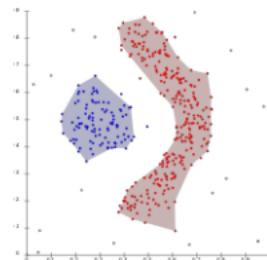
Broad Machine Learning Tasks



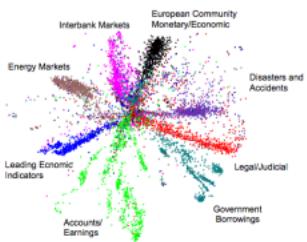
Classification



Regression



Clustering



Dimensionality Reduction

spam
detection

face
recognition

...

Classification: Spam Detection

bConnected
powered by Google

in:spam

Click here to enable desktop notifications for UC Berkeley Mail. [Learn more](#) [Hide](#)

Mail ▾

COMPOSE

Inbox

Starred

Sent Mail

Drafts

All Mail

Less ▾

Important

Chats

Spam (78)

Shusen +

in:spam

Delete all spam messages now (messages that have been in Spam more than 30 days will be deleted)

<input type="checkbox"/>	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
<input type="checkbox"/>	Amazon.com	Dr. Brown's Designed To... - We have a recommendation for you Amazon.com Your Amazon
<input type="checkbox"/>	Western Dental & Orthodo.	🎃 We Are Open Today, Celebrate SmileFest - Logo 866-469-5955 Come celebrate with us c
<input type="checkbox"/>	LinkedIn Salary	Shusen, see salaries for professionals like you. - LinkedIn Shusen Wang Compare salaries
<input type="checkbox"/>	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
<input type="checkbox"/>	Western Dental & Orthodo.	Happy Halloween, We Are Open October 31 🎃 - Logo 866-469-5955 Come celebrate with u
<input type="checkbox"/>	kqaqy2	Conference to-shusen - P80806PN840XJRF20V8HLFFP0V8V6206088J8P8BX0JX4NFDB4I
<input type="checkbox"/>	IMCIC 2018	2nd CFP - Complexity, Informatics and Cybernetics - Dear Shusen Wang, Please consider
<input type="checkbox"/>	Amazon.com	Amazon customers get 2 FREE audiobooks today. - Get a Prime member exclusive offer: 4
<input type="checkbox"/>	ICPSA	Call for Manuscripts in the field of Probability and Stochastic Analysis - The 3rd Int'l Con
<input type="checkbox"/>	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 manuscripts submitted to the journal. Once accepted, the paper will be published in the current issue immediately.

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

International Journal of English, Literature and Research and
Science (IJERS) ISSN: 2456-198K(C) | 2249-5493 (P)

DOI (CrossRef): 10.32181/IJERS

Thomson Reuters ResearcherID: P-3738-2015

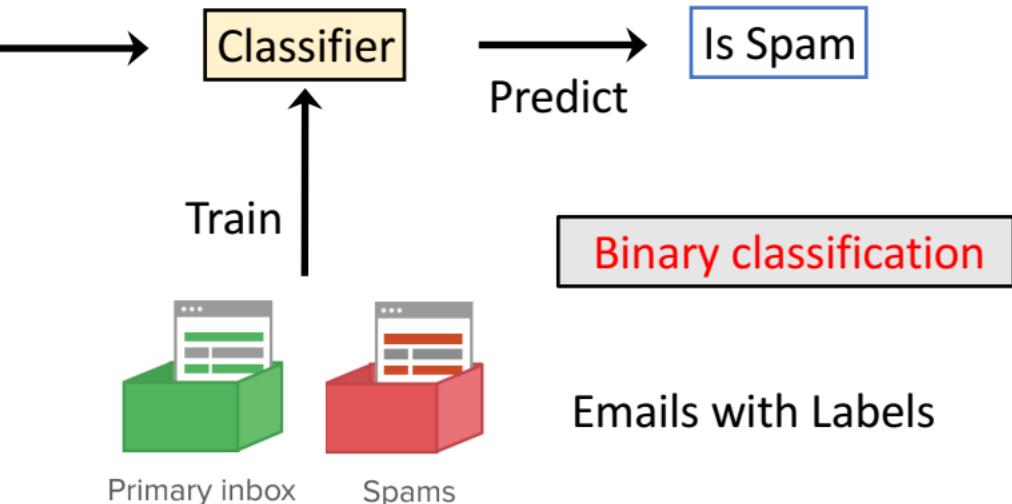
Impact Factor: 4.192, SJIF: 4.072, IBI: 3.2, PIF: 2.465, ISRA-JIF: 1.317,

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

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

International Journal of English, Literature and Science (IJELS)
ISSN: 2456-7820

New Email



Classification: Spam Recognition

The screenshot shows an email inbox interface with several red boxes highlighting specific elements:

- A red box surrounds the "Primary" tab at the top.
- A red box surrounds the "Social" tab at the top.
- A red box surrounds the "Promotions" tab at the top, which is currently selected.
- A red box surrounds the "Spam" folder in the sidebar.

Left Sidebar (Inbox View):

- Inbox: 1,195 (highlighted)
- Starred
- Snoozed
- Sent
- Drafts
- All Mail
- Less
- Important
- Chats
- Spam: 141** (highlighted)
- Trash

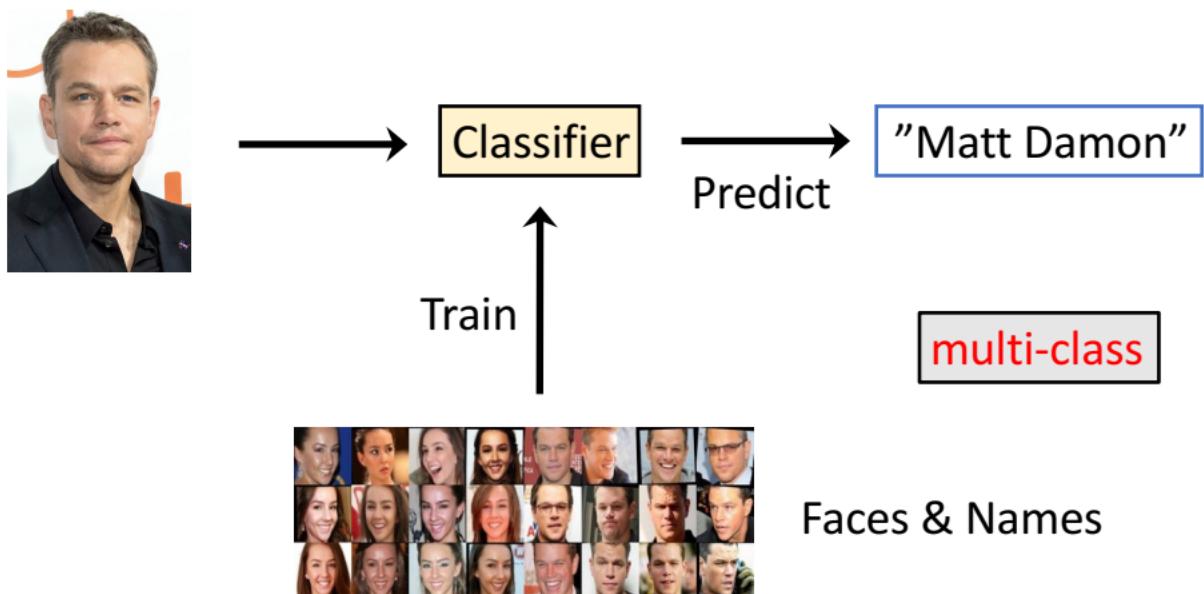
Top Tabs:

- Primary
- Social
- Promotions

Email List:

- Shutterfly: 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



Classification: Image Recognition



Classifier



Predict

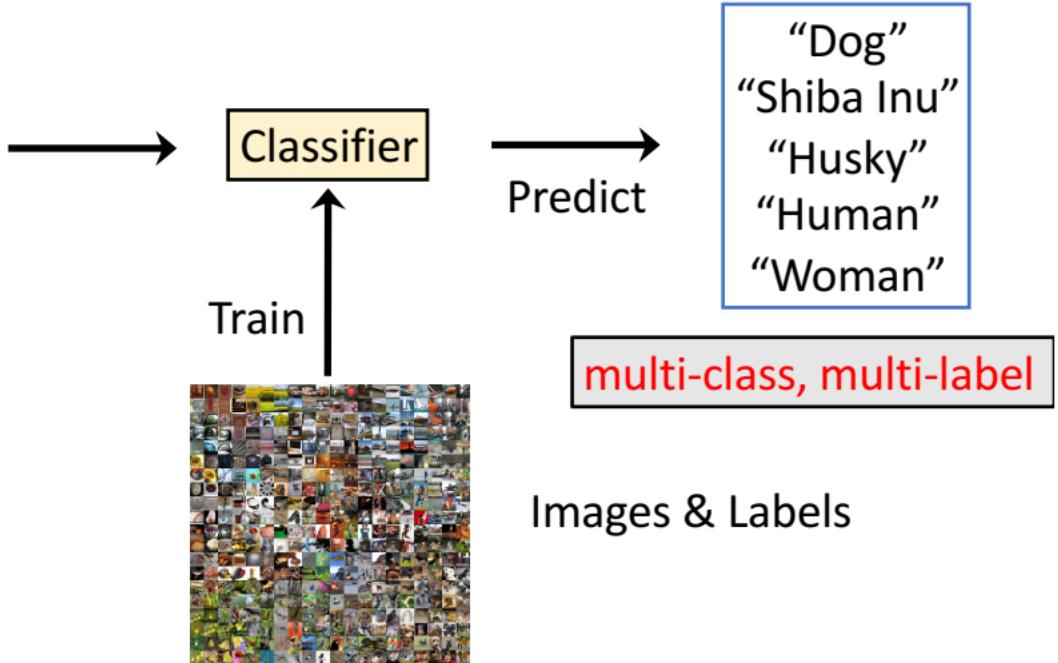
“Dog”
“Puppy”
“Shiba Inu”

multi-class, multi-label

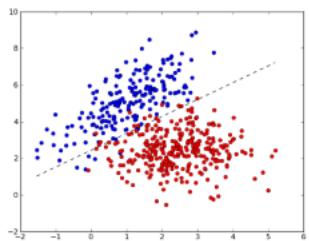


Images & Labels

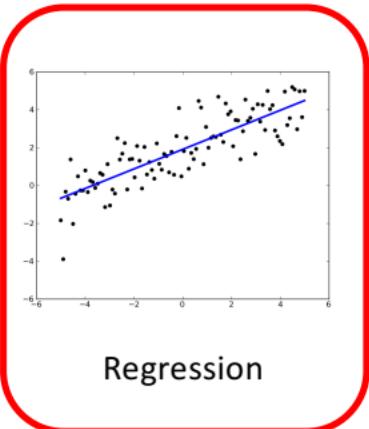
Classification: Image Recognition



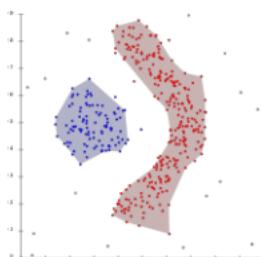
Broad Machine Learning Tasks



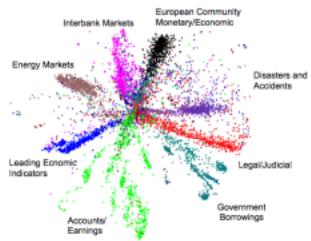
Classification



Regression



Clustering



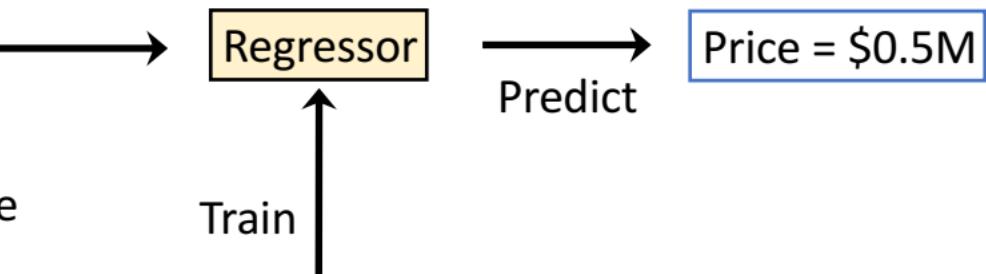
Dimensionality
Reduction

Regression: Housing Price



Features of a House

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:

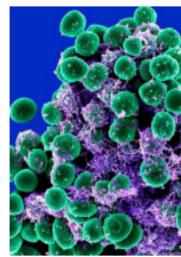
Categories:

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

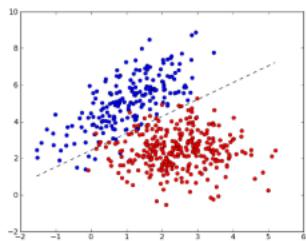
“Class 1” and “Class 4” are categories. They are not ordered!

Class 1

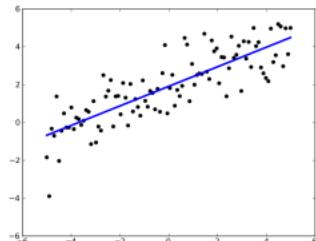


Class 4

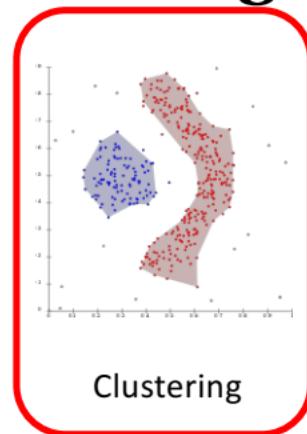
Broad Machine Learning Tasks



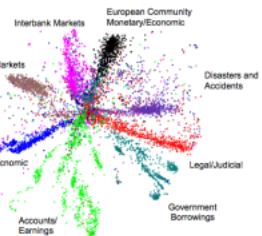
Classification



Regression



Clustering



Dimensionality
Reduction

Supervised Learning

Unsupervised Learning

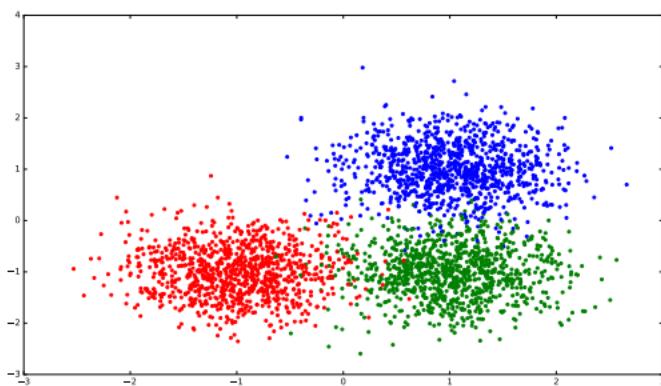
Unsupervised Learning

Supervised Learning

Clustering: Unsupervised Learning

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

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



Clustering: Unsupervised Learning

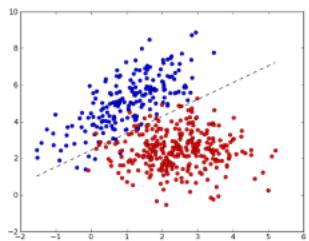
Supervised learning : learn function from **labeled** training data.

- Classification and regression are supervised learning.

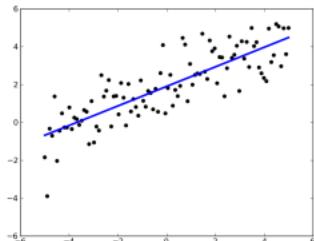
Unsupervised learning : draw inferences from datasets consisting of input data **without label**.

- Clustering is unsupervised learning.

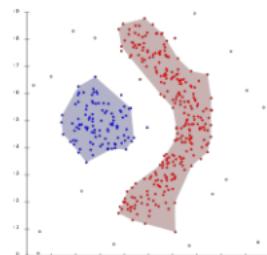
Broad Machine Learning Tasks



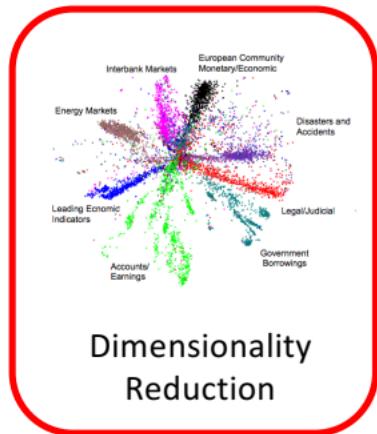
Classification



Regression

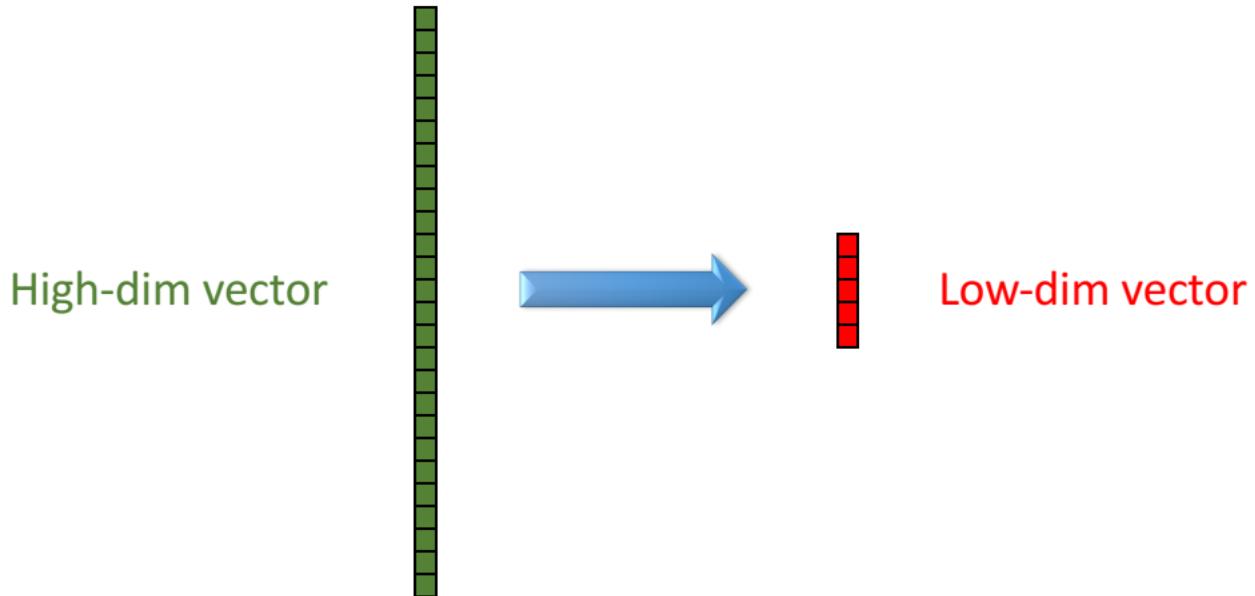


Clustering



Dimensionality
Reduction

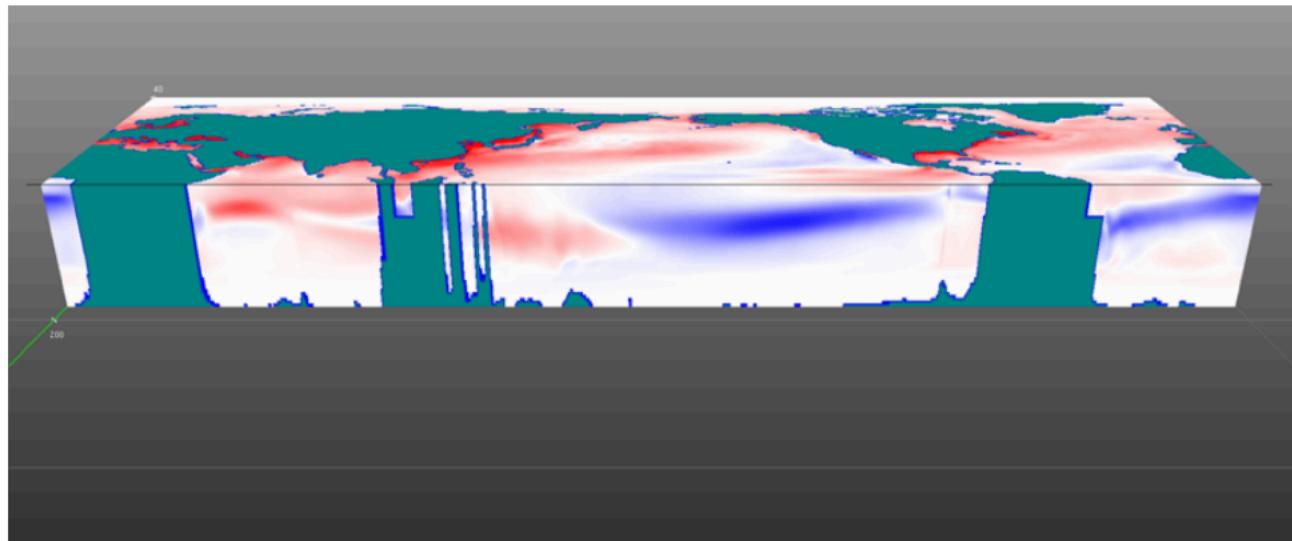
Dimensionality Reduction



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)

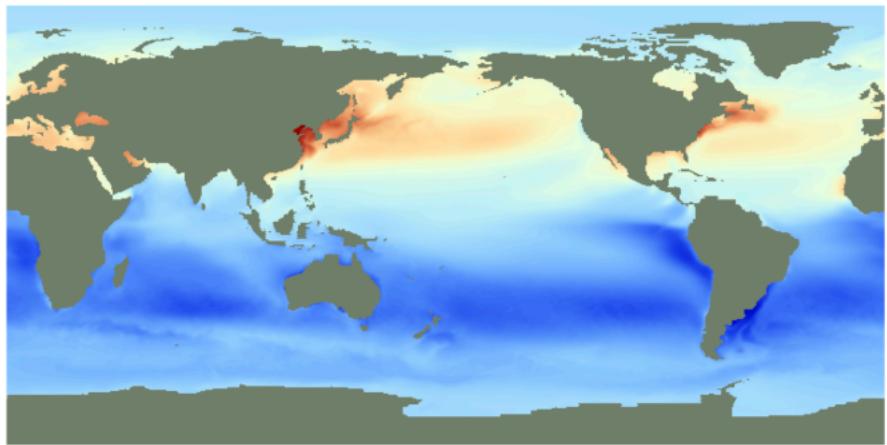
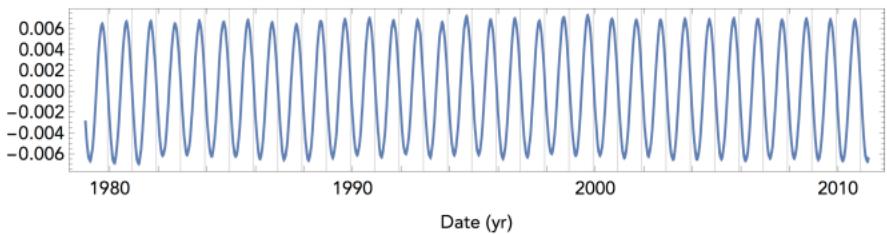
Dimensionality Reduction



ocean temperatures taken 1979—2011 at 6 hour intervals

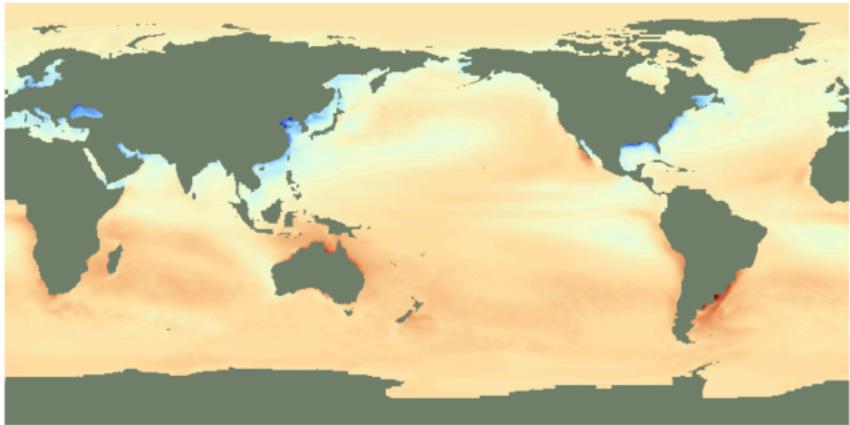
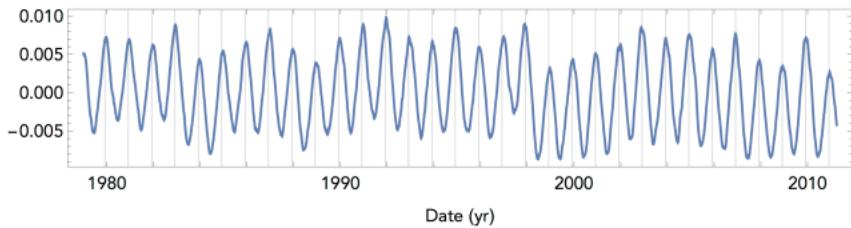
The 1st Principal Components

Annual cycle



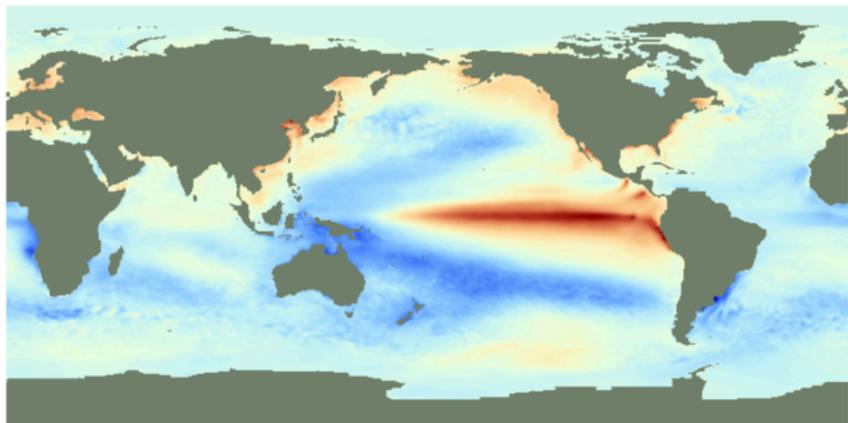
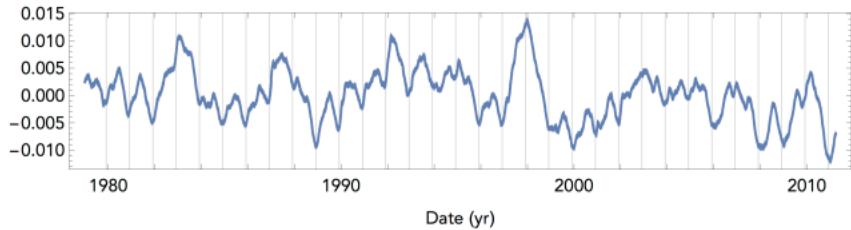
The 2nd Principal Components

Annual cycle



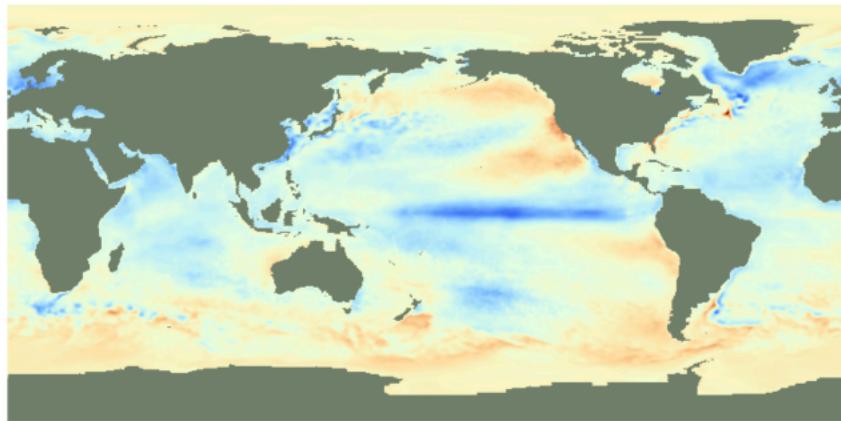
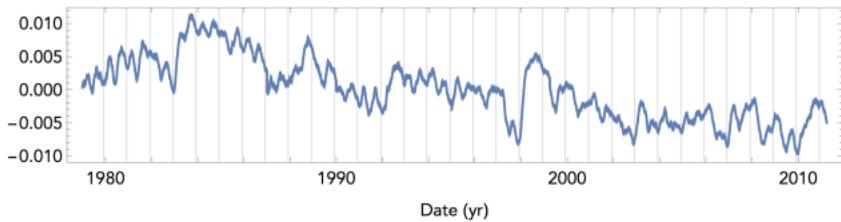
The 3rd Principal Components

El Nino

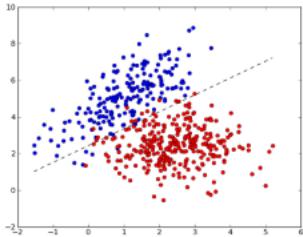


The 4th Principal Components

La Nina



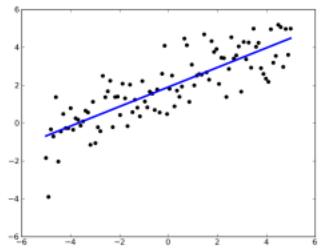
Broad Machine Learning Tasks



Classification



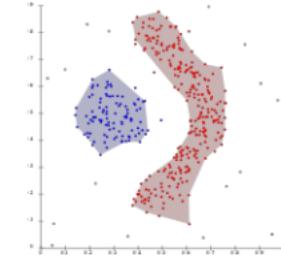
Ranking



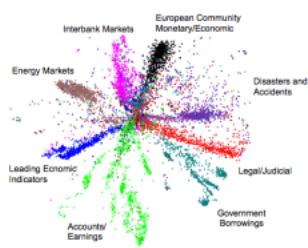
Regression



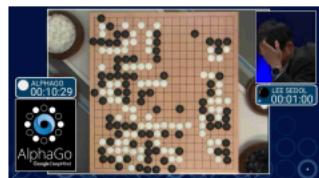
Recommendation



Clustering



Dimensionality Reduction



Reinforcement Learning

Tasks, Methods, & Algorithms

Tasks

Regression

Classification

Clustering

Dim Reduction

Tasks

Regression

Classification

Clustering

Dim Reduction

Methods

Neural Networks

SVM

Logistic Regression

Decision Tree

Nearest Neighbor

Tasks

Regression

Classification

Clustering

Dim Reduction

Methods

Neural Networks

SVM

Logistic Regression

Decision Tree

Nearest Neighbor

Algorithm

Gradient Descent (GD)

Stochastic GD

Coordinate Descent (CD)

Dual CD

Newton's Method