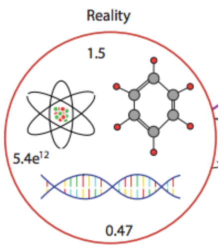
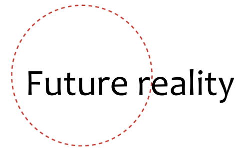
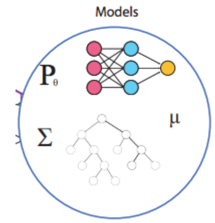






Statistics 154, Spring 2019

Modern Statistical Prediction and Machine Learning

Lecture 18:
Classification

Instructor: Bin Yu
(binyu@berkeley.edu); office hours: Tu: 9:30-10:30 am;
Wed:**1:30-2:30 pm;** office: 409 Evans

GSIs: Yuansi Chen (Mon: 10-12; 4-6); Raaz Dwivedi (Mon: 12-2; 2-4)
yuansi.chen@berkeley.edu; raaz.rsk@berkeley.edu
(Yuansi: Tuesday 1-3; Raaz: Monday 10:30-11:30, Thurs. 9:30-10:30)

Taxonomy

of Machine Learning/Statistics

Labeled Data

Indirect (reward)

Unlabeled Data

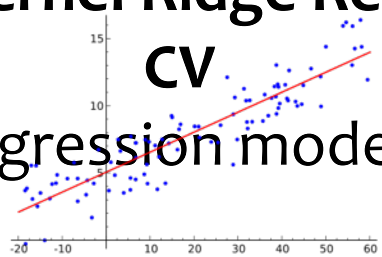
Supervised Learning

Reinforcement & Bandit Learning

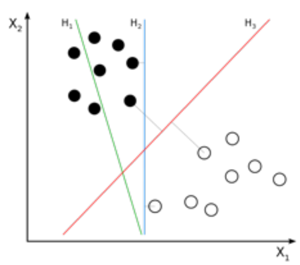
Unsupervised Learning

LS, GD, RR, Lasso, Kernel Ridge Reg CV

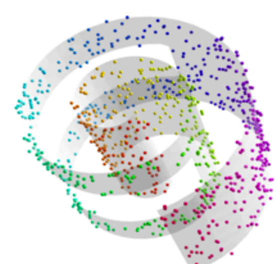
Regression models



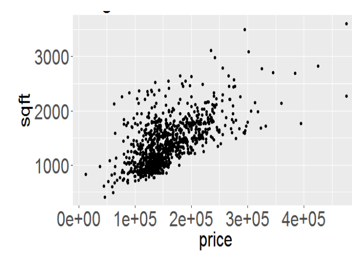
Classification



Dimensionality Reduction



Clustering



Thanks to J. Gonzalez

Example: Predicting credit card default in a 3-circle representation

Current credit card holder data including their card applications, credit reports

? ? ? ? ? ? ? ? ? ?

Prediction rules

? ? ? ? ? ? ? ? ? ?

Q: Are they similar to the current card holders?

Future credit card applicants

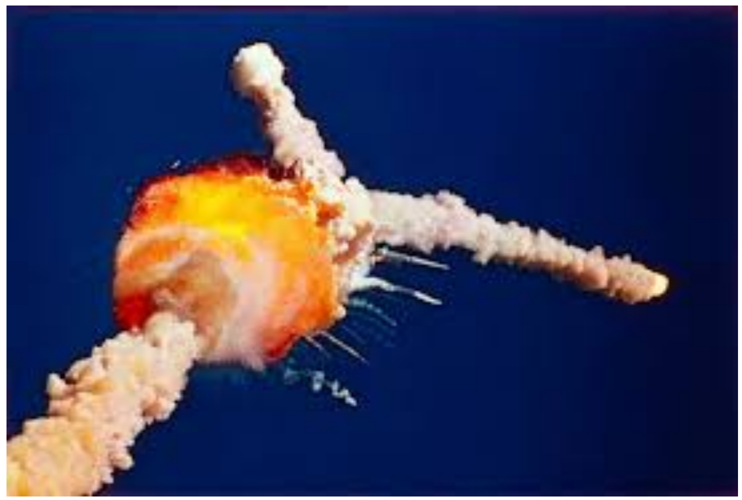
Actions: approval or not

Binary classification – an important supervised learning problem

Generally, each data unit has a feature vector and a label +1 or -1

Examples from class:

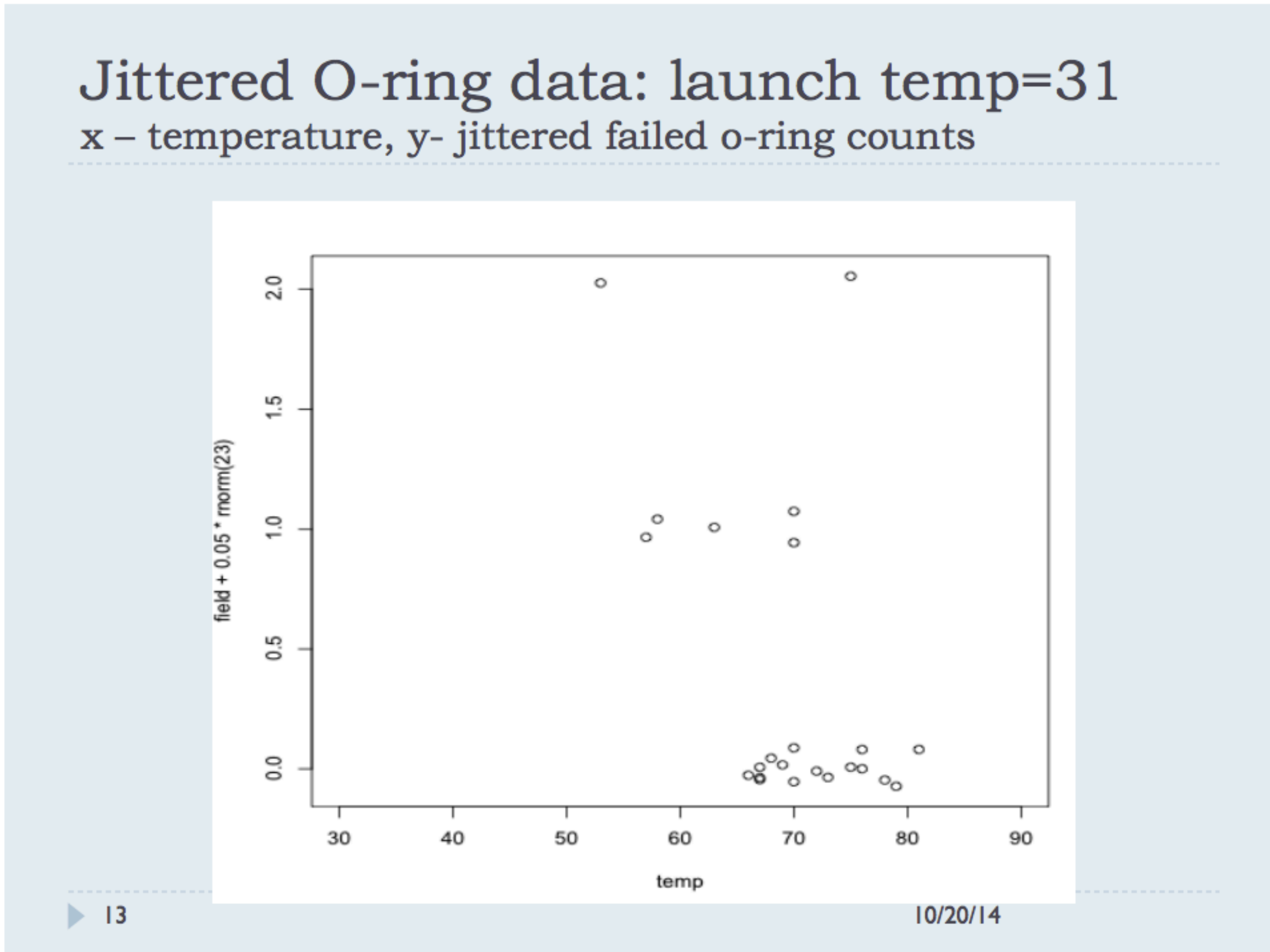
Space Shuttle Challenger Disaster on Jan. 29, 1986



Challenger comes apart after liftoff
(photo by Bruce Weaver of AP)

Statistical recommendations matter!

O-ring failure was identified as the reason after



Data source: R data

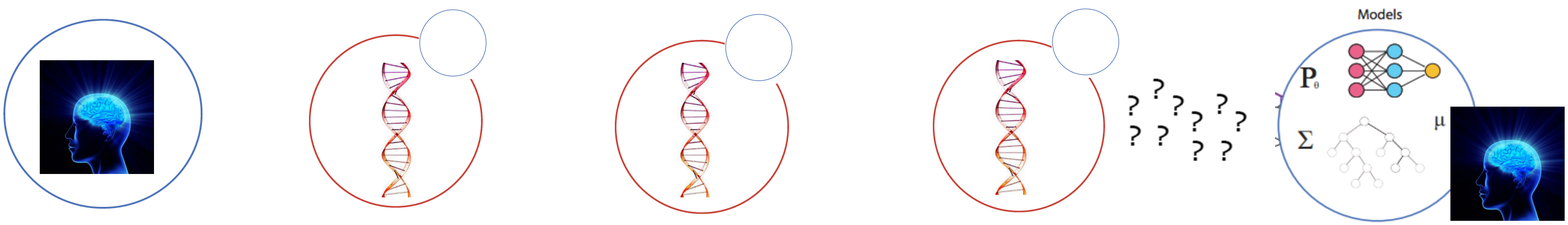
Lesson:
do not do robust analysis automatically

- Outliers could be the most relevant data points for a particular domain problem

- Another example is stress testing in banking industry

- Yet another example is how to price a gold mine in Australia: median or mean?

A data analysis report should try to discuss the whole data science life cycle



Domain Problem → Data collection /storage → **Data cleaning** → Data visualization or EDA → **Unsupervised and supervised ML**

↓

Revised domain problem ← Get new data ← Interpretation data results ← **Model validation** ← Post-hoc analysis/EDA

http://www.odbms.org/2015/04/data-wisdom-for-data-science/

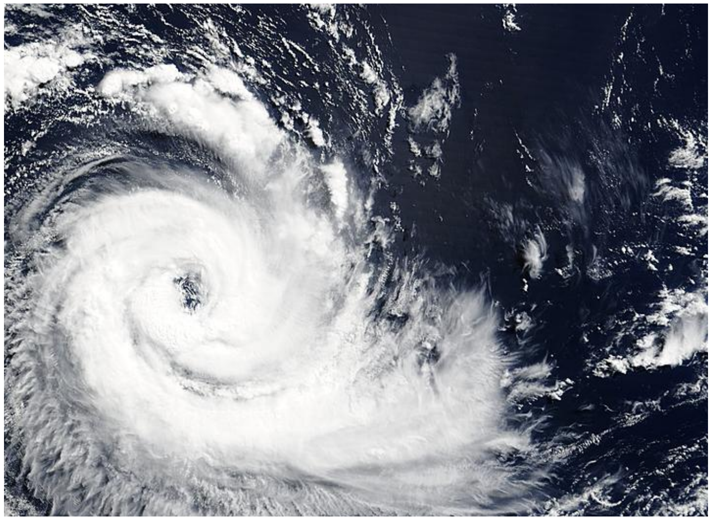
Banking image credit: https://www.kapturecrm.com/banking-crm/

Lab 2: Cloud Detection over Arctic Regions

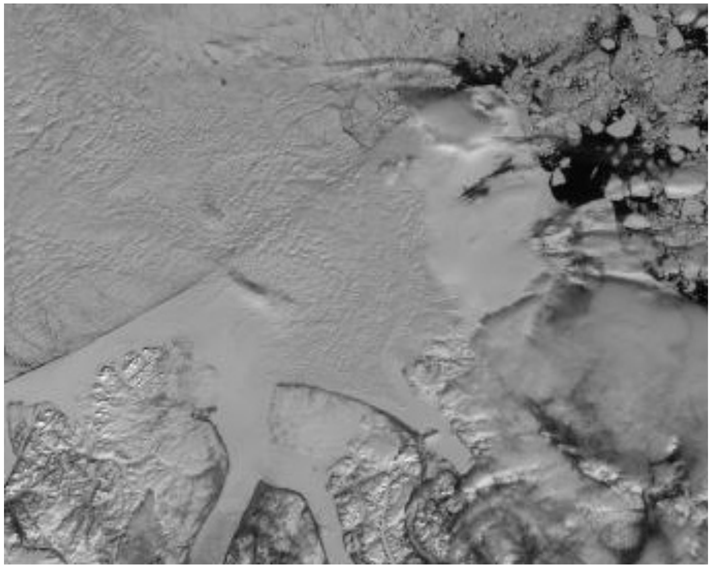
- Uncertainties about cloud radiation feedback on the global climate are among the greatest obstacles in understanding and predicting earth's future climate.

- Clouds above snow- and ice-covered surfaces are especially difficult to detect because their temperature and reflectivity are similar to those of the surface.

Human expert labels are used as "ground truth", but expensive and not available on line.



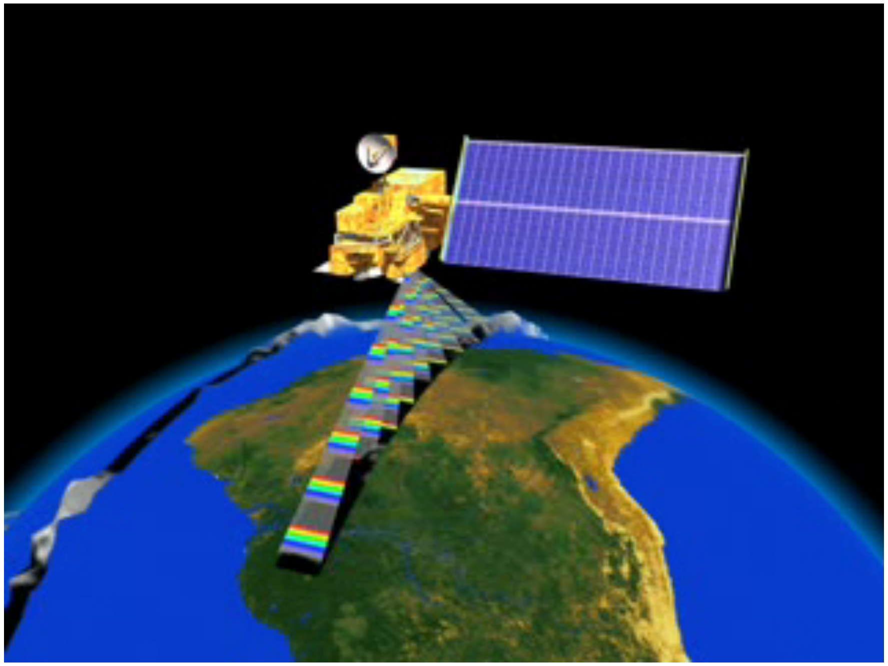
Over Ocean



Over Ice and Snow

Algorithms

Cloud Detection Based on MISR Images



MISR has 9 angles

0^{O} *(AN)*,

$\pm$ 26.1^{O} *(AF, AA)*,

$\pm$ 45.6^{O} *(BF, BA)*,

$\pm$ 60^{O} *(CF, CA)*,

$\pm$ 70.5^{O} *(DF, DA)*

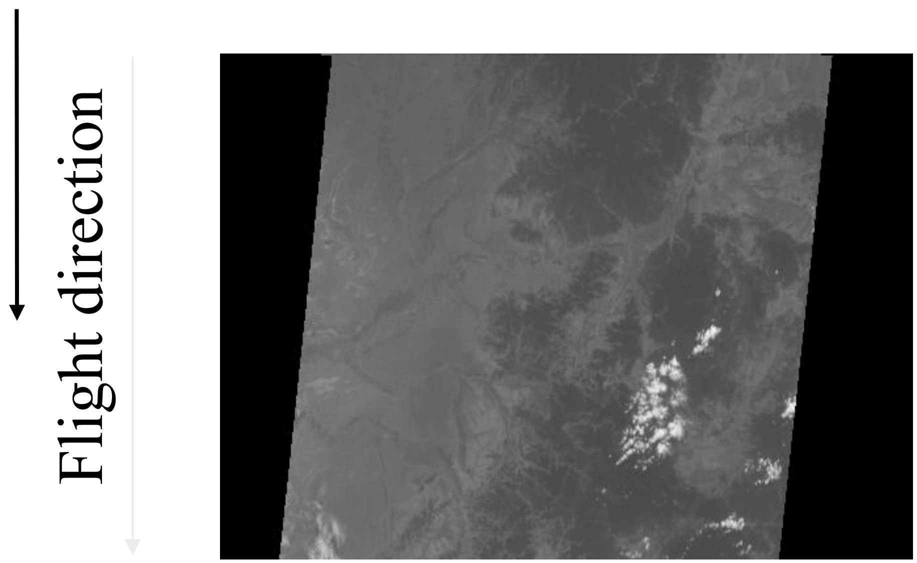
- **Multi-angle Imaging Spectre Radiometer (MISR) was launched by NASA on December 18, 1999.**
- **Built and maintained for NASA by the Jet Propulsion Laboratory (*JPL*) in Pasadena, California.**

- 4 wavelengths in each angle.
  (443nm, 555nm, 670nm, and 865nm Near Infrared Red)

Challenges

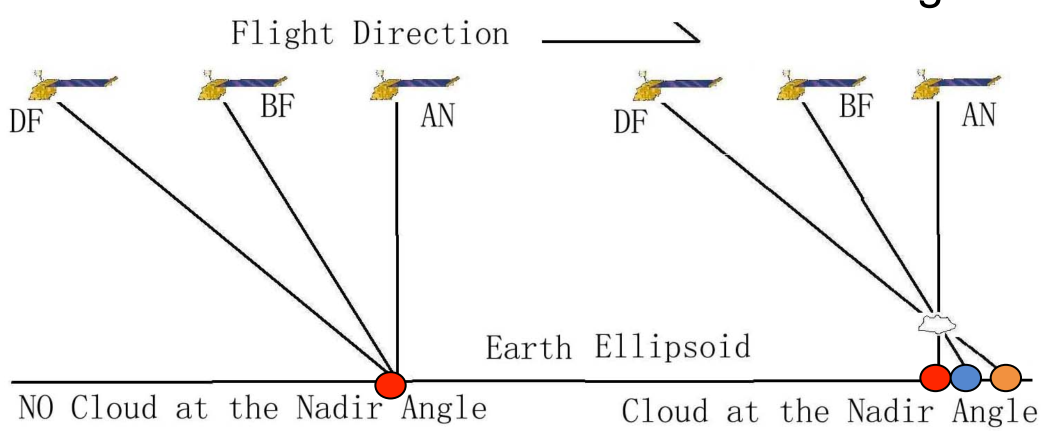
- Organization, transmission, and visualization of these massive data (MISR: 3.3 megabits/s on average and 9.0 megabits/s peak time)
- Streaming data or online processing
- Data fusion among EOS data sources

MISR Operational Cloud Detection Algorithm



Clouds are registered at different ground locations in different angles.

The MISR operational SDCM retrieves the cloud height and cloud movement by matching the same clouds in three angles.



The algorithm works well over dark surfaces, such as deep ocean and vegetation covered land surface, but does not work well over snow and ice covered surfaces because good matching is very difficult.

Lab 2: data

For each pixel:

- Raw measurements (4 channels and each 9 angles)
- Three features – through discovery data analysis and domain knowledge (via interactions with the MISR science team)

Plan

- LDA and QDA (used in the cloud detection project)
- Support Vector Machines (SVMs) – kernel trick is used
- Logistic Regression

Data analysis demo: digits classification using the MNIST data set

LDA (Linear Discriminate Analysis) and QDA (Quadratic Discriminate Analysis)

- LDA – decision boundaries are linear

- QDA – decision boundaries are quadratic

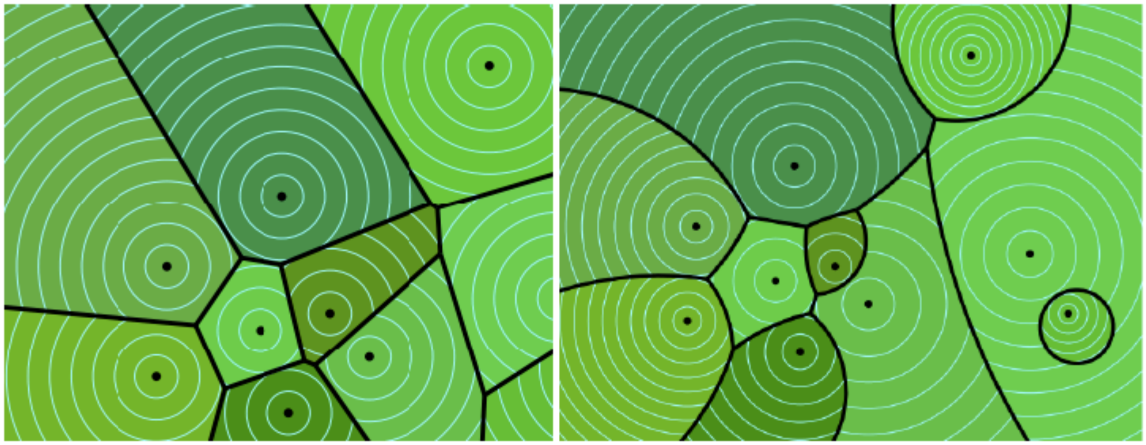


Figure 3: LDA (left) vs QDA (right): a collection of linear vs quadratic level set boundaries. Source: Professor Shewchuk's notes

Ideas behind LDA and QDA

- Model the distribution of the features for each class in a parametric manner
- Estimate the parameters in the distributions
- Classify according to which class gives the highest probability of the data point (or a feature vector)

How do LDA and QDA relate to data?

Current credit card holder data including their card applications, credit reports

? ? ? ? ? ? ? ? ? ?

LDA or QDA models about data generation
i.i.d. is assumed

? ? ? ? ? ? ? ? ? ?

Q: Are they similar to the current card holders?

Future credit card applicants

Actions: approval or not

When the model assumptions are wrong

- LDA and QDA can be viewed as prediction algorithms (without the probabilistic interpretations)

- They can be evaluated in terms of prediction error on future data

- They can also be evaluated in terms of usefulness to downstream goals (e.g. how effective they are to help climate predictions)

Math behind LDA and QDA

- Blackboard derivations based on CS 189 (Spring 2018) Notes 18 by Prof. Anant Sahai

Reading assignment

- Notes from CS189 Spring 2018 by Prof. Anant Sahai – n18.pdf