Introduction to Machine Learning

Bilal Ahmed

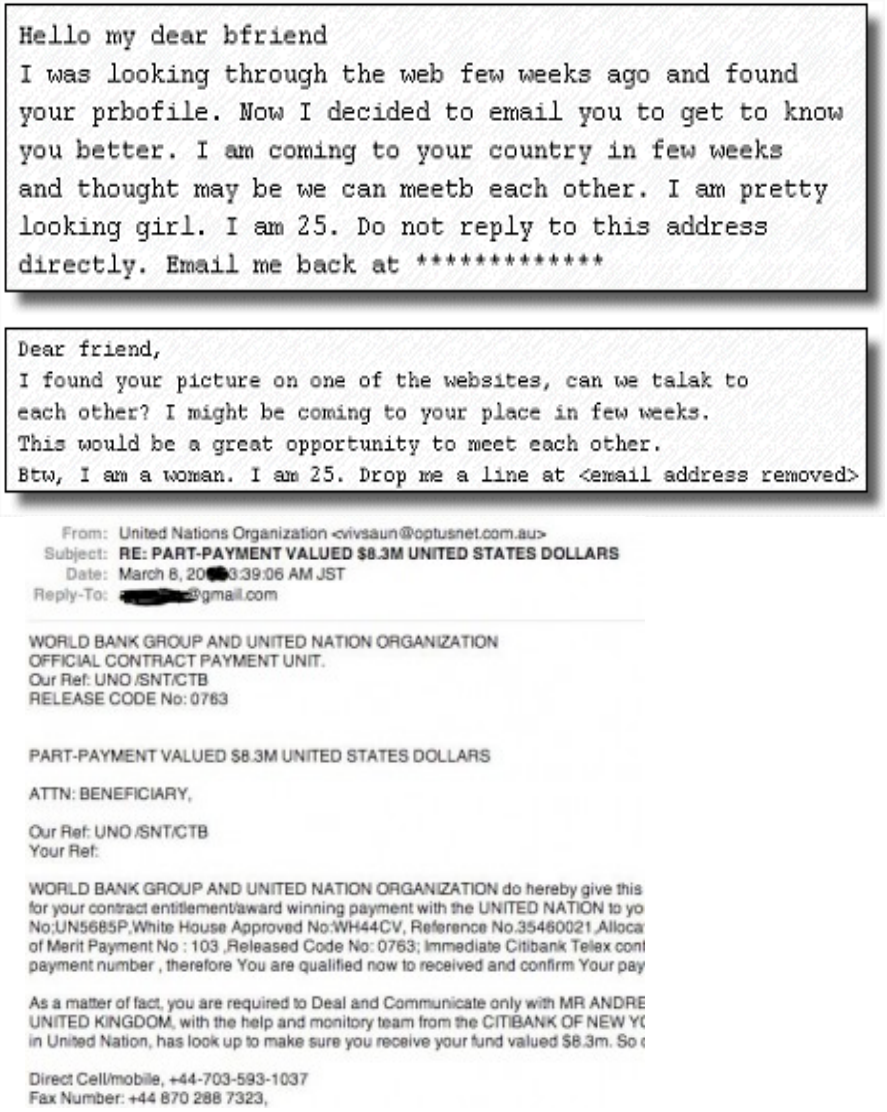
01/12/2015

What is Machine Learning?

- Algorithms that "learn" from data.
  - What do they learn?
  - How do they learn?

- A lot of our decisions/judgements/actions are based on data
  - Medical diagnosis
  - Should I drop this course?

Application Domains



Hello my dear bfriend
I was looking through the web few weeks ago and found your prbofile. Now I decided to email you to get to know you better. I am coming to your country in few weeks and thought may be we can meetb each other. I am pretty looking girl. I am 25. Do not reply to this address directly. Email me back at **********

Dear friend,
I found your picture on one of the websites, can we talak to each other? I might be coming to your place in few weeks. This would be a great opportunity to meet each other. Btw, I am a woman. I am 25. Drop me a line at <email address removed>

From: United Nations Organization <vivsaun@optusnet.com.au>
Subject: RE: PART-PAYMENT VALUED $8.3M UNITED STATES DOLLARS
Date: March 8, 20[redacted]3:39:06 AM JST
Reply-To: [redacted]@gmail.com

WORLD BANK GROUP AND UNITED NATION ORGANIZATION
OFFICIAL CONTRACT PAYMENT UNIT.
Our Ref: UNO /SNT/CTB
RELEASE CODE No: 0763

PART-PAYMENT VALUED $8.3M UNITED STATES DOLLARS

ATTN: BENEFICIARY,

Our Ref: UNO /SNT/CTB
Your Ref:

WORLD BANK GROUP AND UNITED NATION ORGANIZATION do hereby give this
for your contract entitlement/award winning payment with the UNITED NATION to yo
No:UN5685P,White House Approved No:WH44CV, Reference No.35460021,Alloca
of Merit Payment No : 103 ,Released Code No: 0763; Immediate Citibank Telex conf
payment number , therefore You are qualified now to received and confirm Your pay

As a matter of fact, you are required to Deal and Communicate only with MR ANDRE
UNITED KINGDOM, with the help and monitory team from the CITIBANK OF NEW YO
in United Nation, has look up to make sure you receive your fund valued $8.3m. So

Direct Cell/mobile, +44-703-593-1037
Fax Number: +44 870 288 7323,

- Detecting Spam:
  - **Data:** keywords from user marked emails
  - **Task:** automatically decide on spam or ham
  - **Performance Measure:** Accuracy

Application Domains

- OCR **(optical character recognition)**
  - **Data:** Handwritten zip codes
  - **Task:** Identify zip code
  - **Performance Measure:** Accuracy

Application Domains

Output/Target: Discrete

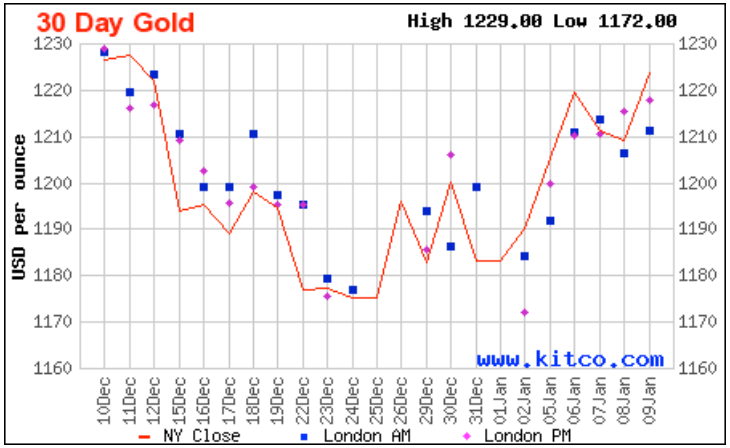
ML Task: Classification

For Spam Detection: Binary Classification

For OCR: Multiclass classification (10 classes)

Training Data: Examples + labels

Application Domains



- Stock Market:
  - **Data:** Trends over the past hour/days/weeks.
  - **Task:** predict stock value for next min/hr/day
  - **Performance Measure:** Deviation from realized value

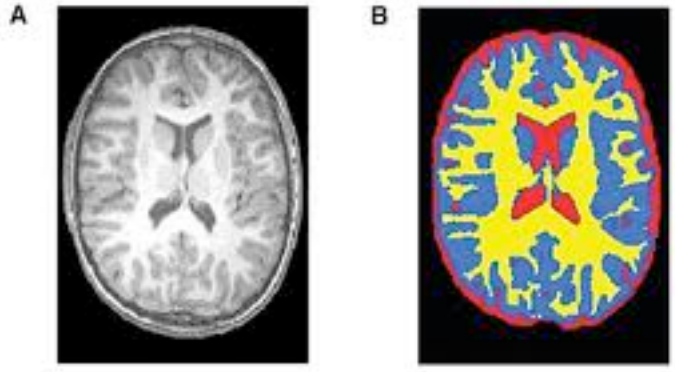
Application Domains

Output/Target: Continuous

ML Task: Regression

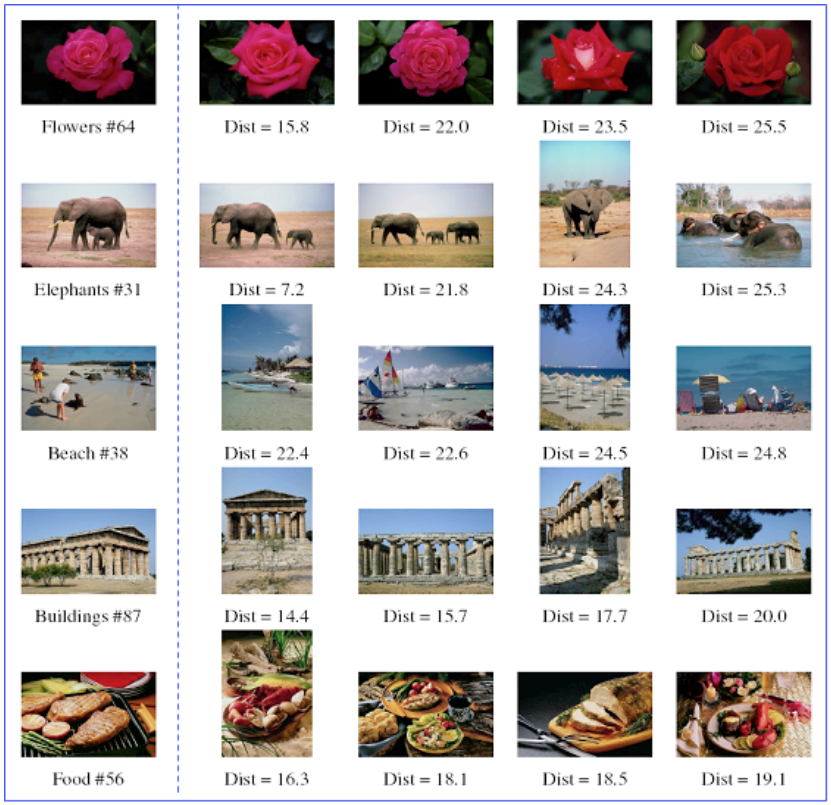
Training Data: Examples + labels (continuous)

Application Domains



- Image Segmentation
  - **Data:** Image
  - **Task:** group pixels that are similar
  - **Performance Measure:** Precision, F-Measure

Application Domains



- Clustering Images for Retrieval
  - **Data:** Images
  - **Task:** group similar images into groups
  - **Performance Measure:** Precision, Recall

Application Domains

Output/Target: Discrete

ML Task: Clustering

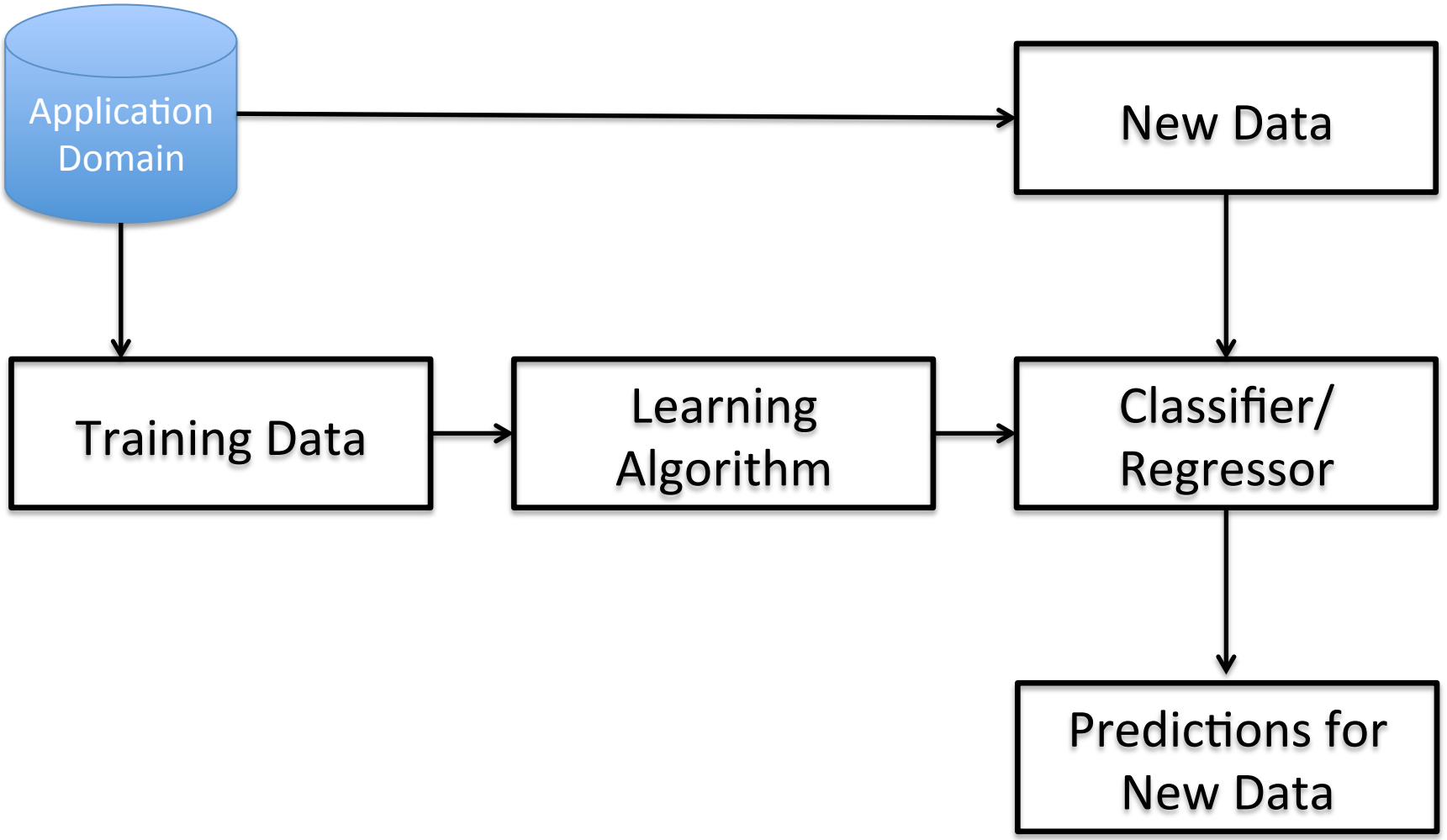
Training Data: Examples only

For image segmentation just pixels

Application Domains

- Other ML tasks:
  - Collaborative Filtering: Recommender systems (Amazon, Netflix, etc.)
  - Ranking: retrieving objects relevant to a query and assigning them a rank. (Google web results, content based image retrieval, etc.)

Machine Learning



Slide taken from Roni Khardon's slides.

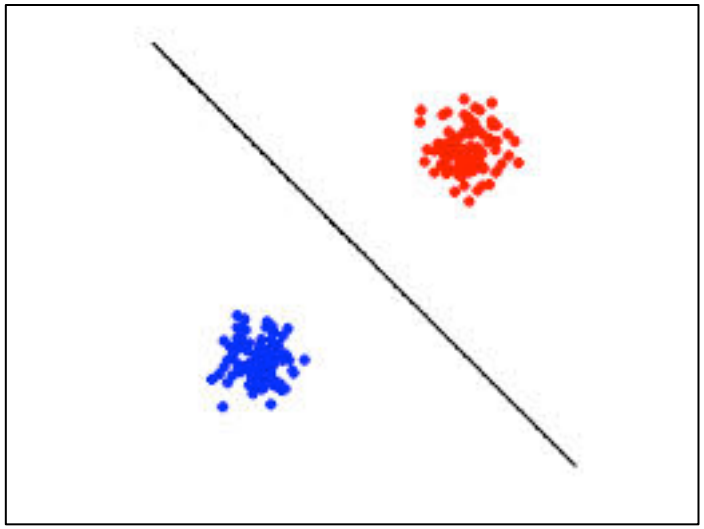
Machine Learning

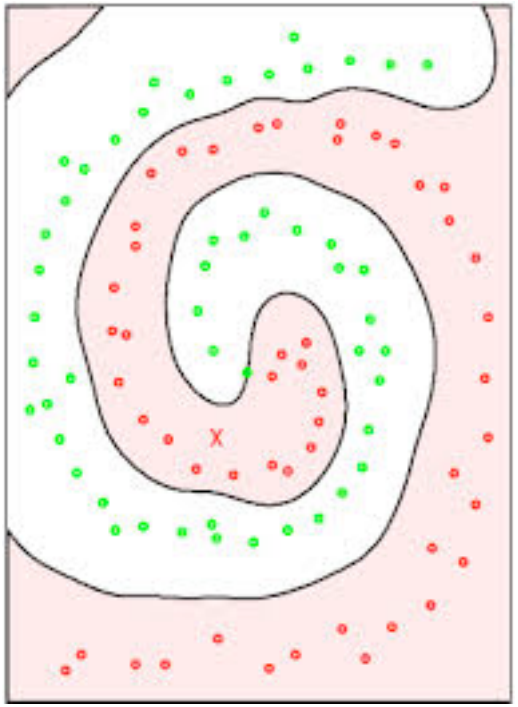
- How do we represent that training/new data?
  - It depends on the application domain
  - and also on our interpretation of the domain

- Digit Recognition:
  - Image represented as a vector (vector data)
  - So all training data can represented as table with each column representing an individual pixel.

Machine Learning

- How do we represent a classifier/regressor?
  - This also depends on the task, its complexity and our judgment





Machine Learning

- Learning Algorithm:
  - How do we analyze/process the training data to create a classifier/regressor that has high "accuracy"?
  - Depends on the type of classifier/regressor we are using.

Slide adapted from Roni Khardon's slides.

Supervised Learning

- Given a training set:
  - Examples: x_i
  - Labels: y_i

- Find a function *f(x)* : X→Y that "correctly" predicts labels for examples.