Machine Learning in Action

Heli Helskyaho

Oracle Code 2018

Miracle Finland Oy

Challenges in Big Data

* More and more data (volume)
* Different data models and formats (variety)
* Loading in progress while data exploration going on (velocity)
* Not all data is reliable (veracity)
* We do not know what we are looking for (value, viability, variability)
* Must support also non-technical users (journalists, investors, politicians,…) (visualization)
* All must be done *efficiently and fast and as much as possibly by machines*

ML in short

* Use the right *Features*
  * with right Algorithms
    * to build the right *Models*
      * that archieve the right *Tasks*

Two types of Methods (techniques)

- Unsupervised learning
  - finds hidden patterns or intrinsic structures in input data
- Supervised learning
  - trains a model on known input and output data to predict future outputs

Miracle Finland Oy

Unsupervised Learning

* Learning from unlabeled input data by finding hidden patterns or intrinsic structures in that data
* Machine learning algorithms find natural patterns in data to make better decisions and predictions possible
* used typically when you
  * don’t have a specific goal
  * are not sure what information the data contains
  * want to reduce the features of your data as a preprocessing for supervised learning

Data for Unsupervised Learning

	A	B	C	D	E	F	G	H	I
46529	2007,1,16,2,1712,1715,1810,1815,WN,990,N252,58,60,45,-5,-3,SJC,BUR,296,3,10,0,,0,0,0,0,0,0								
46530	2007,1,16,2,1228,1230,1327,1330,WN,1191,N374SW,59,60,46,-3,-2,SJC,BUR,296,2,11,0,,0,0,0,0,0,0,0								
46531	2007,1,16,2,907,905,1003,1005,WN,1445,N409,56,60,46,-2,2,SJC,BUR,296,1,9,0,,0,0,0,0,0,0								
46532	2007,1,16,2,1944,1940,2040,2040,WN,1449,N311,56,60,46,0,4,SJC,BUR,296,3,7,0,,0,0,0,0,0,0,0								
46533	2007,1,16,2,650,650,749,750,WN,1650,N364,59,60,46,-1,0,SJC,BUR,296,2,11,0,,0,0,0,0,0,0								
46534	2007,1,16,2,2052,2050,2151,2150,WN,2206,N356,59,60,49,1,2,SJC,BUR,296,2,8,0,,0,0,0,0,0,0,0								
46535	2007,1,16,2,2053,2055,2204,2215,WN,889,N234,71,80,59,-11,-2,SJC,LAS,386,4,8,0,,0,0,0,0,0,0,0								
46536	2007,1,16,2,926,925,1047,1045,WN,1088,N340,81,80,67,2,1,SJC,LAS,386,4,10,0,,0,0,0,0,0,0								
46537	2007,1,16,2,1748,1750,1902,1910,WN,1113,N423,74,80,63,-8,-2,SJC,LAS,386,2,9,0,,0,0,0,0,0,0,0								
46538	2007,1,16,2,2127,2130,2241,2250,WN,1232,N326,74,80,62,-9,-3,SJC,LAS,386,3,9,0,,0,0,0,0,0,0,0								
46539	2007,1,16,2,700,700,816,820,WN,1325,N725,76,80,61,-4,0,SJC,LAS,386,3,12,0,,0,0,0,0,0,0								
46540	2007,1,16,2,1344,1345,1502,1505,WN,2331,N241,78,80,65,-3,-1,SJC,LAS,386,3,10,0,,0,0,0,0,0,0,0								
46541	2007,1,16,2,1552,1555,1709,1715,WN,2583,N236,77,80,64,-6,-3,SJC,LAS,386,4,9,0,,0,0,0,0,0,0,0								
46542	2007,1,16,2,647,635,753,745,WN,123,N659SW,66,70,51,8,12,SJC,LAX,308,7,8,0,,0,0,0,0,0,0								
46543	2007,1,16,2,1833,1835,1936,1945,WN,196,N365,63,70,49,-9,-2,SJC,LAX,308,4,10,0,,0,0,0,0,0,0								
46544	2007,1,16,2,1420,1325,1531,1435,WN,197,N306SW,71,70,52,56,55,SJC,LAX,308,4,15,0,,0,0,0,0,1,0,55								
46545	2007,1,16,2,1652,1650,1800,1800,WN,756,N631SW,68,70,53,0,2,SJC,LAX,308,7,8,0,,0,0,0,0,0,0,0								
46546	2007,1,16,2,755,755,902,905,WN,1247,N642WN,67,70,52,-3,0,SJC,LAX,308,5,10,0,,0,0,0,0,0,0,0								
46547	2007,1,16,2,1619,1620,1727,1730,WN,1577,N628SW,68,70,52,-3,-1,SJC,LAX,308,5,11,0,,0,0,0,0,0,0,0								
46548	2007,1,16,2,1527,1525,1635,1635,WN,1581,N365,68,70,50,0,2,SJC,LAX,308,5,13,0,,0,0,0,0,0,0								
46549	2007,1,16,2,2116,2120,2228,2230,WN,1635,N317SW,72,70,51,-2,-4,SJC,LAX,308,5,16,0,,0,0,0,0,0,0,0								
46550	2007,1,16,2,1429,1430,1535,1540,WN,1664,N619SW,66,70,49,-5,-1,SJC,LAX,308,5,12,0,,0,0,0,0,0,0,0								
46551	2007,1,16,2,1255,1255,1359,1405,WN,1843,N225,64,70,51,-6,0,SJC,LAX,308,3,10,0,,0,0,0,0,0,0,0								
46552	2007,1,16,2,909,910,1040,1025,WN,2087,N684,91,75,50,15,-1,SJC,LAX,308,12,29,0,,0,0,0,0,15,0,0								
46553	2007,1,16,2,1008,955,1116,1105,WN,2164,N601WN,68,70,51,11,13,SJC,LAX,308,6,11,0,,0,0,0,0,0,0,0								
46554	2007,1,16,2,1101,1105,1211,1215,WN,2607,N625SW,70,70,55,-4,-4,SJC,LAX,308,5,10,0,,0,0,0,0,0,0,0								

MiRACLE
Miracle Finland Oy

Clustering

* *Clustering* is the most common method for unsupervised learning and used for *exploratory data analysis to find hidden patterns or groupings in data.*
* *Clustering algorithms*
  * *Hard clustering*
    * each data point belongs to *only one* cluster
  * *Soft clustering*
    * each data point can belong to *more than one* cluster

Supervised Learning

* Learning from known, labelled data
* Training a model on known input and output data to predict future outputs (remember that uncertainty is always involved)

Miracle Finland Oy

Data for Supervised Learning

1	Year,Month,DayofMonth,DayOfWeek,DepTime,CRSDepTime,ArrTime,CRSArrTime,UniqueCarrier,FlightNum,TailNum,ActualElapsedTime,CRSElapsedTime,AirTime,ArrDelay,DepDelay,Origin,Dest,Distance,TaxiIn,TaxiOut,Cancelled,CancellationCode,Diverted,C
2	2007,1,1,1,1232,1225,1341,1340,WN,2891,N351,69,75,54,1,7,SMF,ONT,389,4,11,0,,0,0,0,0,0,0
3	2007,1,1,1,1918,1905,2043,2035,WN,462,N370,85,90,74,8,13,SMF,PDX,479,5,6,0,,0,0,0,0,0,0
4	2007,1,1,1,2206,2130,2334,2300,WN,1229,N685,88,90,73,34,36,SMF,PDX,479,6,9,0,,0,3,0,0,0,31
5	2007,1,1,1,1230,1200,1356,1330,WN,1355,N364,86,90,75,26,30,SMF,PDX,479,3,8,0,,0,23,0,0,0,3
6	2007,1,1,1,831,830,957,1000,WN,2278,N480,86,90,74,-3,1,SMF,PDX,479,3,9,0,,0,0,0,0,0,0
7	2007,1,1,1,1430,1420,1553,1550,WN,2386,N611SW,83,90,74,3,10,SMF,PDX,479,2,7,0,,0,0,0,0,0,0
8	2007,1,1,1,1936,1840,2217,2130,WN,409,N482,101,110,89,47,56,SMF,PHX,647,5,7,0,,0,46,0,0,0,1
9	2007,1,1,1,944,935,1223,1225,WN,1131,N749SW,99,110,86,-2,9,SMF,PHX,647,4,9,0,,0,0,0,0,0,0
10	2007,1,1,1,1537,1450,1819,1735,WN,1212,N451,102,105,90,44,47,SMF,PHX,647,5,7,0,,0,20,0,0,0,24
11	2007,1,1,1,1318,1315,1603,1610,WN,2456,N630WN,105,115,92,-7,3,SMF,PHX,647,5,8,0,,0,0,0,0,0,0
12	2007,1,1,1,836,835,1119,1130,WN,2575,N493,103,115,88,-11,1,SMF,PHX,647,7,8,0,,0,0,0,0,0,0
13	2007,1,1,1,2047,1955,2332,2240,WN,2608,N733SW,105,105,89,52,52,SMF,PHX,647,7,9,0,,0,49,0,0,0,3
14	2007,1,1,1,2128,2035,2245,2200,WN,139,N348,77,85,66,45,53,SMF,SAN,480,3,8,0,,0,0,0,3,0,42
15	2007,1,1,1,935,940,1048,1105,WN,747,N358,73,85,63,-17,-5,SMF,SAN,480,2,8,0,,0,0,0,0,0,0
16	2007,1,1,1,1251,1245,1405,1410,WN,933,N413,74,85,65,-5,6,SMF,SAN,480,2,7,0,,0,0,0,0,0,0
17	2007,1,1,1,1729,1645,1843,1810,WN,1054,N416,74,85,64,33,44,SMF,SAN,480,3,7,0,,0,3,0,0,0,30
18	2007,1,1,1,825,825,941,950,WN,1106,N383SW,76,85,63,-9,0,SMF,SAN,480,3,10,0,,0,0,0,0,0,0
19	2007,1,1,1,1042,1040,1158,1205,WN,1554,N316SW,76,85,66,-7,2,SMF,SAN,480,2,8,0,,0,0,0,0,0,0
20	2007,1,1,1,1726,1725,1839,1850,WN,1604,N691WN,73,85,63,-11,1,SMF,SAN,480,3,7,0,,0,0,0,0,0,0
21	2007,1,1,1,1849,1820,2016,1940,WN,1975,N308SW,87,80,69,36,29,SMF,SAN,480,3,15,0,,0,20,0,7,0,9
22	2007,1,1,1,2219,2105,2332,2225,WN,2083,N205,73,80,62,67,74,SMF,SAN,480,3,8,0,,0,0,0,0,0,67
23	2007,1,1,1,2012,1940,2131,2105,WN,2577,N603SW,79,85,66,26,32,SMF,SAN,480,3,10,0,,0,9,0,0,0,17

A process of supervised learning 1/2

1. Train
    1. Load data
    2. Pre-process data
    3. Learn using a method and an algorithm
    4. Create a model
    * iterate until you find the best model

A process of supervised learning 2/2

2. Predict (use the model with new data)

1. New data
2. Pre-process data
3. Use the model
4. Get predictions
5. Integrate the models into applications

Supervised Learning, methods/techniques

* Predictive models
  * Classification
  * Regression

Supervised Learning, Classification

* Classification models are trained to *classify* data into *categories*.
* They predict discrete responses
  * an email is genuine or spam
  * a tumor is small, medium size, or large
  * a tumor is cancerous or benign
  * a person is creditworthy or not
* For example applications like medical imaging, speech recognition, and credit scoring

Supervised Learning, Classification

* Can the data be tagged or categorized? Can it be separated into specific groups or classes?
  * Classification might be the right answer

Supervised Learning, Regression

* To predict continuous responses
  * changes in temperature
  * fluctuations in electricity demand
* For example applications like forecasting stock prices, handwriting recognition, acoustic signal processing, failure prediction in hardware, and electricity load forecasting.

MIRACLE Miracle Finland Oy

Approximation!

* ML always gives an approximated answer
* Some are better than others, some are useful

* search for patterns and trends
* Prediction accuracy: the higher the number the better it will work on new data
* several models, choose the best, but still: all approximations! There is no correct answer…

Real life use cases for ML

* Online shopping (Amazon, Search, recommendations)
* Voice-to-Text, Smart Personal Assistants (mobile services: ”recipe for bread”, ”find the nearest grocery”)
  * Siri, Google Assistant, Alexa, Echo, Cortana,…
* Facebook
* …

Facebook, some use cases for ML, the Products

* News Feed ranking
* Ads
* Search
* Sigma
* Lumos
* Facer
* Language Translation
* Speech Recognition

News Feed

* ML is used for
  * ranking and personalizing News Feed stories
  * filtering out offensive content
  * highlighting trending topics
  * ranking search results, and much more.
* *General models* are *trained* to determine various user and environmental factors that should ultimately determine the rank order of content.
* *The model* is used to generate a *personalized* set of the best posts, images, and other content to display from thousands of candidates, and the best ordering of this chosen content.

Miracle Finland Oy

Ads

* Online advertising allows advertisers to only bid and pay for measurable user responses, such as clicks on ads.
  * As a consequence, click prediction systems are central to most online advertising systems.
* General Ads models are trained to learn *how user traits, user context, previous interactions, and advertisement attributes* can be most predictive of the likelihood of clicking on an ad, visiting a website, and/or purchasing a product.
* Inputs are run through a trained model to immediately determine which ads to display to a particular Facebook user.

Predicting the Clicks

* The click prediction system needs to be robust and adaptive, and capable of learning from *massive* volumes of data.
* At Facebook they use a model which *combines decision trees with logistic regression*
* Based on their experience: the most important thing is to have *the right features* (those capturing historical information about the user or ad dominate other types of features) and the right model
* Measures: the accuracy of prediction

Miracle Finland Oy

Search

* Launches a *series* of distinct and specialized *sub-searches* to the various verticals, e.g., videos, photos, people, events, etc.
* A *classifier* layer is run atop the various search verticals *to predict which of the many verticals to search* (searching all possible verticals would be inefficient)
* The classifier and these search verticals consist of
  * an *offline* stage to *train* the models
  * and an *online* stage to *run the models* and perform the classification and search

Miracle Finland Oy

Sigma

* General *classification and anomaly detection framework* that is used for a variety of internal applications (site integrity, spam detection, payments, registration, unauthorized employee access, and event recommendations)
* Sigma includes *hundreds of distinct models running in production everyday*
  * each model is trained to detect anomalies (e.g. classify content)

Lumos

* *Extract* high-level *attributes* and *embeddings* from *an image* and its *content*
  * That data can be *used as input* to other products and services
    * for example as it were text.

Facer

* Facebook's *face detection and recognition framework*
* Given an image
  * *finds* all of the faces in that image
  * *runs a user-specific* facial-recognition algorithm to determine the likelihood of that face belonging to one of your top-N friends who have enabled face recognition
* This allows Facebook to suggest which of your friends you might want to tag within the photos you upload.

Language Translation

* Service that manages *internationalization* of Facebook content
* Supports *translations* for more than 45 languages (as the source or target language)
  * supports more than 2000 translation directions
  * serves 4.5B translated post impressions every day
* Each language pair direction has its own model
  * multi-language models are being considered

Miracle Finland Oy

Speech Recognition

* Converts *audio streams into text*
* Provides automated captioning for video
* Most streams are English language
  * other languages will be available in future
* Additionally, non-language audio events are also detected with a similar system (simpler model).

Miracle Finland Oy

FBLearner Flow

* The platform consists of three core components:
  * *an authorship and execution environment* for custom distributed workflows
  * *an experimentation management UI* for launching experiments and viewing results
  * *numerous predefined pipelines for training* the most commonly used machine learning algorithms at Facebook.

FBLearner Predictor

* Facebook's *internal inference engine* that uses the models trained in FBLearner Flow *to provide predictions* in real time.
  * Can be used as a multitenancy service
  * or as a library that can be integrated in product specific backend services
  * Is used by multiple product teams at Facebook, many of which require low latency solutions.
* The direct integration between Flow and Predictor also helps with
  * running online experiments
  * managing multiple versions of models in productions

Frameworks for deep learning

* Two distinct but synergistic frameworks for deep learning at Facebook:
  * PyTorch, which is optimized for *research*
  * Caffe2, which is optimized for *production*

PyTorch

* PyTorch is the framework for AI *research* at Facebook which enables rapid experimentation
  * Flexibility
  * Debugging
  * Dynamic neural networks
* *Not* optimized for production and mobile deployments (Python)
* When research projects produce valuable results, *the models need to be transferred to production.*
  * Traditionally, rewriting the training pipeline in a product environment with other frameworks.

MiRACLE
Miracle Finland Oy

Caffe2

* Facebook's in-house *production* framework
  * For training and deploying large-scale machine learning models
* Focuses on several key features required by products:
  * Performance
  * cross-platform support
  * coverage for fundamental machine learning algorithms (convolutional neural networks (CNNs), recurrent networks (RNNs), and multi-layer perceptrons (MLPs)) and up to tens of billions of parameters

Miracle Finland Oy

Facebook

* “we noticed that the largest improvements in accuracy often came from quick experiments, feature engineering, and model tuning rather than applying fundamentally different algorithms”
* An engineer may need to attempt hundreds of experiments before finding a successful new feature or set of hyperparameters.

Miracle Finland Oy