Real-world examples/scenarios for Machine Learning techniques

|  |  |
| --- | --- |
| **Week** | **Topics** |
| 1 | Overview of Machine Learning: history, relation to classical statistics **Text:** Chapter 1; |

Activity: Distinguish between a human and a computer by asking questions and analyzing the answers.

The game is played as follows.

There are four actors: we will call them Gina, George, Herb and Connie (the first letter of the names will help you remember their roles). The teacher coordinates proceedings. The rest of the class forms the audience. Gina and George are *go-betweens*, Herb and Connie will be answering questions. Herb will give a human’s answers, while Connie is going to pretend to be a computer. The class’s goal is to find out which of the two is pretending to be a computer and which is human. Gina and George are there to ensure fair play: they relay questions to Herb and Connie but don’t let anyone else know which is which. Herb and Connie are in separate rooms from each other and from the audience.

See template/format (search on google) day\_19\_sec

Use resources from here:

<http://www.cs4fn.org/machinelearning/>

**Week 2. Linear Regression.**

* **House Price Prediction**. Linear regression analysis can help a builder to predict how much houses it would sell in the coming months and at what price.
* **Predict Economic Growth**. Economists use Linear Regression to predict the economic growth of a country or state.
* **Score Prediction**. Sports analyst use linear regression to predict the number of runs or goals a player would score in the coming matches based on previous performances.

**Week 3. Online Learning Algorithms**

* **Online Email Categorization**. Consider an email categorization task in which the instances are email messages. We cast the online categorization problem as a label ranking problem. Label ranking is the task of ordering labels with respect to their relevance to an input instance.
* **Time series Prediction** using Online Learning Algorithms (examples of time series: stock values, economic variables, weather: e.g., local/global temperature, earthquakes, energy demand, signal processing, sales forecasting, etc.)

**Week 4. Support Vector Machines**

* **Face detection** – SVMs classify parts of the image as a face and non-face and create a square boundary around the face.
* **Text and hypertext categorization** – SVMs allow Text and hypertext categorization for both inductive and transductive models. They use training data to classify documents into different categories. It categorizes on the basis of the score generated and then compares with the threshold value.
* **Classification of images** – Use of SVMs provides better search accuracy for image classification. It provides better accuracy in comparison to the traditional query-based searching techniques.
* **Bioinformatics** – It includes protein classification and cancer classification. We use SVM for identifying the classification of genes, patients on the basis of genes and other biological problems.
  + **Protein fold and remote homology detection** – Apply SVM algorithms for protein remote homology detection.
* **Handwriting recognition** – We use SVMs to recognize handwritten characters used widely.

**Week 5. Decision Trees. Boosting**

* **Selecting a flight to travel.** Suppose you need to select a flight for your next travel. How do we go about it? We check first if the flight is available on that day or not. If it is not available, we will look for some other date but if it is available then we look for may be the duration of the flight. If we want to have only direct flights then we look whether the price of that flight is in your pre-defined budget or not. If it is too expensive, we look at some other flights else we book it!
* **Predicting Library Book Use.** Forecasting book usage helps librarians to select low-usage titles and move them to relatively distant and less expensive off-site locations that use efficient compact storage techniques. *(Reference****:***[*Silverstein, C., & Shieber, S. M. (1996). Predicting individual book use for off-site storage using decision trees. The Library Quarterly, 66(3), 266-293*](https://scholar.google.com/scholar?hl=en&as_sdt=0%2C33&q=Silverstein%2C+C.+and+Shieber%2C+S.+M.+Predicting+individual+book+use+for+off-site+storage+using+decision+trees.+Lib.+Q.+66%283%29%3A266%E2%80%93293%2C+1996.&btnG=)*.)*
* **Cancer cell classification.** Boosted trees are most useful when you have unbalanced class distribution, e.g. cancer vs non-cancerous cell classification where cancerous cells are rare say 1%.

*(another possible format that I could use)*

**Week 6 and 7. Deep Learning.**

* Restore colors in B&W photos and videos.
  + ["Let there be color!"](http://dl.acm.org/citation.cfm?id=2925974) is a computer system that can automatically restore colors in B&W photos. You can read more about it [here](http://hi.cs.waseda.ac.jp/~iizuka/projects/colorization/en/) and see plenty other examples [here](http://hi.cs.waseda.ac.jp/~iizuka/projects/colorization/extra.html).
* Automatically Adding Sounds to Silent Movies.
  + The system must synthesize sounds to match a silent video. The system is trained using 1000 examples of video with sound of a drum stick striking different surfaces and creating different sounds. A deep learning model associates the video frames with a database of pre-rerecorded sounds in order to select a sound to play that best matches what is happening in the scene. The system was then evaluated using a turing-test like setup where humans had to determine which video had the real or the fake (synthesized) sounds. Youtube example [here](https://www.youtube.com/watch?time_continue=9&v=0FW99AQmMc8).
* Self-driving cars.
  + Everybody heard about this one, and now you can actually see them in action. In this video [a Tesla electric vehicle](https://www.tesla.com/) drives without human intervention. Notice how it distinguishes different type of objects, including people and road signs.
* Robotics.
  + Deep Learning is also heavily used in robotics these days. This is a field of itself which I won't get into, but at least two examples of my favorite robots by [BostonDynamics](http://www.bostondynamics.com/): SpotMini and [Atlas](http://www.bostondynamics.com/robot_Atlas.html). The robots react to people pushing them around, they also get up when falling, and can even take care of pretty elaborate tasks that require gentle and care, like unloading a dish washer.
* Voice Recognition.
  + [Google released WaveNet](https://deepmind.com/blog/wavenet-generative-model-raw-audio/) and [Baidu released Deep Speech](https://www.youtube.com/watch?v=kAnJdvf_KeE), both are Deep Learning networks that generated voice automatically. You may ask what's the big deal? [Siri](https://www.wikiwand.com/en/Siri) and [Alexa](https://www.wikiwand.com/en/Amazon_Alexa) can talk as well. To date, text2voice systems were not completely autonomous in the way they created new voices, they were (manually) trained to do so. The systems created today learn to mimic human voices by themselves and improve with time. When letting an audience try to differentiate them from a real human speaking, it is much harder to do so. While we are not there yet in terms of automatic voice generation, Deep Learning is taking us a step closer to giving computers the ability to speak like humans do.

Week 8. Clustering

* **Biology**: classification of plants and animals given their features;
* **City-planning**: identifying groups of houses according to their house type, value and geographical location;
* **Earthquake studies**: clustering observed earthquake epicenters to identify dangerous zones;
* **Insurance**: identifying groups of motor insurance policy holders with a high average claim cost;
* **Marketing**: finding groups of customers with similar behavior given a large database of customer data containing their properties and past buying records;
* **WWW**: document classification; clustering weblog data to discover groups of similar access patterns.

[Source](http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/). Another **excessive list** of [applications of cluster analysis](http://en.wikipedia.org/wiki/Cluster_analysis#Applications) from **Wikipedia**

**Week 9. Expectation Maximization (EM).**

*Estimate missing values or learn from incomplete data.*

* EM is frequently used for data clustering (see examples from week 8)
* Find parameters of a hidden Markov model (HMM) (i.e., the Baum-Welch algorithm, later in week 13)
* **Manage risk of a portfolio.** With the ability to deal with missing data and observe unidentified variables, EM is becoming a useful tool to price and manage risk of a portfolio.
* **Medical image reconstruction**. The EM algorithm (and its faster variant ordered subset expectation maximization) is also widely used in medical image reconstruction, especially in positron emission tomography and single photon emission computed tomography.
* **Asset return prediction.** Consider asset returns as a sequence of states or regimes. Each regime is characterized by its own descriptive statistics including mean and volatility. Example regimes could include low-volatility and high-volatility. We can also assume that asset returns will transition between these regimes based on probability. By framing the problem this way we can use mixture models, which are designed to try to estimate the sequence of regimes, each regime’s mean and variance, and the transition probabilities between regimes. The most common is the Gaussian mixture model (GMM). The underlying model assumption is that each regime is generated by a Gaussian process with parameters we can estimate. Under the hood, GMM employs an expectation-maximization algorithm to estimate regime parameters and the most likely sequence of regimes.

**Week 10. Dimensionality reduction**

* Customer relationship management.
* Text categorization.
* Image retrieval.
* Gene expression microarray data analysis
* Face recognition
* Handwritten digit recognition.
* **Email Classification.**
  + Given a database of emails, classify (using some machine learning numerical algorithm) each email as spam/not spam. To achieve this goal, you construct a mathematical representation of each email as a bag-of-words vector. Not all dimensions (words) of your vectors are informative for the spam/not spam classification. For instance, words “lottery”, “credit”, “pay” would be better features for spam classification than “dog”, “cat”, “tree”. For a mathematical way to reduce dimension we will use PCA.

**Week 11 and 12 Graphical Techniques**

* **Manufacturing.** Graphical Models has its applications in Manufacturing field. Making the production of low cost and most reliable components at a high rate is possible. If all components of a production system (i.e. machine tool operation, dispatching etc) work on optimized parameters. Compute this by using graphical models.
* **Finance.** Graphs, because they are pictures. They are particularly appropriate for presentation of financial information. Presenting financial information requires a careful understanding of both “what you want to say” and “who you need to say it to“.
* **Steel Production.** To calculate the emission of carbon dioxide for Steel’s deoxidation we use Graph model.
* **Handwriting Recognition.** We can use Graphical models to recognize hand writing. It can also use in several applications. It uses hand writing for identification.
* **Telecommunication Network Diagnosis.** We use Graphical models for diagnosing issues in it. It is used for help and to resolve them.
* **Object Recognition in Images.** Graphical models provide a powerful framework for encoding. It provides the statistical structure of visual scenes. It also provides developing corresponding learning and inference algorithms.

**Week 13. Structured Prediction**

|  |  |
| --- | --- |
| **Weekly Schedule** | |
| **Week** | **Topics** |
| 1 | Overview of Machine Learning: history, relation to classical statistics **Text:** Chapter 1; |
| 2 | **Linear Regression:** model specification issues, maximum likelihood estimation, ridge regression **Text:** Chapter 7, except 7.4 and 7.6 |
| 3 | **Logistic Regression:** model fitting methods (steepest descent, Newton’s, L1 and L2 regularization, etc.) **Online learning algorithms**: structured prediction, regret minimization, stochastic optimization, risk minimization, LMS algorithm, perceptron algorithm, relation to Bayesian techniques. **Text:** Chapter 8.1-8.3.6, 8.5, 13.3, |
| 4 | **Support Vector Machines:** uses in regression and classification, optimization, choice of parameter C, probabilistic interpretation **Kernel Methods:** RBF kernels, Mercer kernels, Matern kernels, Linear kernels, String kernels, Kernels derived from probabilistic generative models; the “kernel trick” and its uses: nearest neighbor classification, K-medoids clustering, ridge regression, and PCA **Text:** Chapter 14.1, 14.2, 14.5, supplemental notes by Ng |
| 5 | **Decision Trees:** basis in information theory, when appropriate to use, growing and pruning, detecting and avoiding over-fitting; pros and cons with respect to other ML techniques; random forests; ID3 and related algorithms; use of bias; training with incomplete data **Boosting:** why it works so well, specific types of boosting, ensemble learning **Text:** Chapters 2.8, 16.2, 16.4, 16.6 , supplemental notes by Mitchell, Yoav & Schapire |
| 6 | **Deep Learning:** background in feed-forward neural networks (multilayer perceptrons) and Restricted Boltzmann machines; deep generative models (sigmoid networks, Boltzmann machines, belief networks) **Text:** Chapters 16.5, 27.7, 28.1-28.2 |
| 7 | **Deep Learning (continued):** training of deep networks, applications of deep networks **Midterm Review** **Text:** Chapters 28.3-28.4 |
| 8 | **Midterm Exam Clustering:** measures of dissimilarity; k-means clustering, clustering in mixture models **Text:** Chapters 25.1, 11.1-11.3 |
| 9 | **Expectation Maximization:** basic idea, Jensen’s inequality, the EM algorithm, EM applied to various distributions (mixture of Gaussians), theoretical basis **Text:** Chapters 11.4, supplemental notes by Ng |
| 10 | **Dimensionality Reduction:** classical Principal Components Analysis, PCA theorem; Singular Value Decomposition, Probabilistic PCA, EM algorithm for PCA, PCA for categorical data, PCA for paired and multi-view data **Text:** Chapters 12.2, 12.4, 12.5 |
| 11 | **Graphical Techniques:** Bayesian nets, Directed Gaussian Models (DGM), plate notation, d-separation, Markov random fields, Hammersley-Clifford theorem and its effects, training maxent models, pseudo-likelihood, stochastic maximum likelihood learning **Text:** Chapter 10, 19.1-19.5 |
| 12 | **Graphical Techniques (continued):** belief propagation (serial, parallel, Gaussian, and other variants), variable elimination, junction trees, intractability of exact inference, approximate inference methods, max sum and max product **Text:** Chapter 20 |
| 13 | **Structured Prediction:** Markov and Hidden Markov (HMM) models, inference algorithms (forward, Viterbi, backward filtering, etc.), training methods, Baum-Welch algorithm, model selection, Conditional Random Fields (CRF), label-bias problem, uses in handwriting recognition, protein analysis, and stereopsis **Text:** Chapters 17.1-17.5, 19.6-19.7, CRF tutorial paper by Sutton & McCallum |