What are these? Data Mining, Machine Learning, Artificial Intelligence, Natural Language Processing, Information Retrieval....

July 26, 2016

Let's first see what the Internet says!

Quora Answers

Data Mining

- Data mining focuses on the discovery of (previously) unknown properties on the data.
- Data Mining: Data Mining is the process of extraction of data or previously unknown data patterns from a large set of data. Hence as the word suggests, you 'mine for specific data' from a large data sets. Thus typical data mining algorithms check for a specific pattern or trend from the given data set. Many times the output of data mining algos are used as input for machine learning algos.
- Data Mining is a cross-disciplinary field that focuses on discovering properties of data sets.
- There are different approaches to discovering properties of data sets. Machine Learning is one of them.

Data Mining

- This field is mostly concerned with extracting information from a vast amount of data. Google Search is one example of such a system. It is not exactly a technical subject rather application of different algorithms related to NLP, Machine Learning and AI. All Search applications, Text Summarization, Question Answering systems (SIRI) etc are example of this.
- Another example is Association Rule Mining/Pattern Mining being applied in Retail Industry to mine buying behavior of consumers from vast amount of historical buying behavior.

Machine Learning

- Machine learning focuses on prediction, based on known properties learned from the training data.
- Machine learning is relates to the designing and development of machine that can learn itself from a given set of data to achieve a desirable outcome without it being explicitly coded. Hence machine learning literally means 'a machine that learns on its own'. Applications: spam mail detection, optical character recognition etc
- This subject deals with turning data in to information and taking decisions based on that.
- Some algorithms are Classification, Clustering, Regression, Feature Selection, etc.

Artificial Intelligence

- This subject deals with Logic, Reasoning, Graph Traversing/Mining etc. It deals with automatic ways of reasoning and reaching to a conclusion by computers.
- Search and Optimization are two big use cases of AI. An example could be Google Maps. It shows you shortest path to reach to your destination. AI is one way to achieve that.
- Robot Navigation, Automatic Clinical Decision System, Knowledge representation in NLP are few other applications.

- What is the difference between Artificial Intelligence, Machine Learning, Statistics, and Data Mining ?
- Would it be accurate to say that they are 4 fields attempting to solve very similar problems but with different approaches?
- What exactly do they have in common and where do they differ?
- If there is some kind of hierarchy between them, what would it be?

- Machine learning is a science that involves development of self-learning algorithms. These algorithms are more generic in nature that it can be applied to various domain related problems.
- Data mining is a practice of applying algorithms (mostly Machine learning algorithms) with the data available from domain to solve domain related problems.
- Statistics is a study of how to collect, organizes, analyze, and interpret numerical information from data. Statistics can slip into two taxonomy called Descriptive statistics and Inferential statistics. Descriptive statistics involves method of organizing, summering and picturing information from data. Inferential statistics invokes method of using information from sample to draw conclusion about the population.

Contd..

- Machine learning uses statistics (mostly inferential statistics) to develop self learning algorithms.
- Data mining uses statistics (mostly Descriptive statistics) on results obtained from algorithms, it used to solve the problem.
- Data mining as a field emerged to solve problems in the miscellaneous domain (particularity in business), acquired different techniques and practices that are used in different field of studies.
- Artificial Intelligence is a science to develop a system or software to mimic human to respond and behave in a circumference. As field with extremely broad scope, AI has defined its goal into multiple chunks. Later each chuck has become a separate field of study to solve its problem.

Time frame

- Statistics – 1749
- Artificial Intelligence – 1940
- Machine leaning – 1946
- Data mining – 1980

An exact question!!

What is the difference between (i) Artificial Intelligence, (ii) Data Mining, (iii) Information Retrieval, (iv) Natural Language Processing, and (v) Machine Learning & Pattern Recognition?

(i) Artificial Intelligence, (ii) Data Mining, (iii) Information Retrieval, (iv) Natural Language Processing, and (v) Machine Learning & Pattern Recognition

They are all related in some way but they differ on what they especialize or how they look at the problems.

- Artificial Intelligence is a classic term that is still in use. It is used to refer to techniques that could be used to replicate intelligence. How to identify intelligence is still debated. After realizing (according to some) that full intelligence is not possible to achieve, the field was broken down in subfields like Learning, Language, Vision, etc, each focusing on a specific area of intelligence
- Data Mining is a way of getting information out of data, but it doesn't actually imply artificial intelligence. You can use AI to do data mining but non-AI techniques are also used.
- Natural Language Processing is a field in itself but some consider it a part of AI. It deals only with processing language, written or spoken, to different ends like Human-Computer Interaction, Data analysis, subtitling and many other applications.

Contd..

- Information Retrieval is more concerned with indexing and classification. Again, you can use AI for this but it's not required. Think of Google's PageRank.
- Machine Learning is the study of techniques (and the techniques themselves) to make computers learn to solve different problems. This techniques depend on the context and can be applied in other fields (like Data Mining, Information Retrieval or Natural Language Processing).
- Pattern Recognition is the automated recognition of patterns, independently of where or how this patterns happen. It usually involves Machine Learning but not necessarily, and the goal is to produce a classifier that given some input it assigns a category. Again, this can be seen as part of AI.