ARTIFICIAL INTELLIGENCE IN INDUSTRY

Bridging Industry-Academia gap

Agenda – part 1

- Who am I?
- Why you should focus for the next 2 hours?
- What is Artificial Intelligence (AI)?
- Elements of AI
- Use of AI in industry – few examples
- Different roles of Data Science

Agenda – Part 2

- Supervised learning and Unsupervised learning
- Typical algorithms of supervised and unsupervised learning
- A brief recap of linear regression
- Mention of some Maths/Stats components which are used in industry on a daily basis
- Challenge problem – discussion of data set
- Solution of the problem
- Homework challenge

Introduction

- Current role – Data Scientist
- Current Designation – Senior Manager
- Total Experience – 14+ years
- Worked in the fields of – Analytics, Architecture, Business Process Modelling, Enterprise Application Integration, Project Management and Pre-sales.
- Core skills - Natural Language Processing, Machine Learning and Watson Analytics.
- Hobby – Sports (cricket, football, tennis)… Active member of IBM corporate team

Reason to Focus

As per the recent statistics, the education sector in India is poised to witness major growth in the years to come as India will have world's largest tertiary-age population and second largest graduate talent pipeline globally by the end of 2020. India's economy is also expected to grow at a fast pace; rapid industrialization would require a gross incremental workforce of ~250 million by 2030; India could potentially emerge as a global supplier of skilled manpower. However, despite these encouraging statistics, a major segment of graduates remain unemployable - according to 'National Employability Report' 2016, which is based on a study of more than 1,50,000 engineering students who graduated in 2015 from over 650 colleges, 80% of them were unemployable and only 3% had suitable skills to be employed in software or product market.

(March 2017 - www.peoplematters.in)

How to tackle the challenge?

- Increase the enrolment ratios in higher education
- Alignment of curriculum with industry requirements
- Emphasis on skill-based education
- Internships that give workplace exposure
- Up-skilling the faculty

Where this lecture is going to help?

Workplace exposure through industry/corporate interactions –

- Well-timed and well-deliberated exposure to the industry provides a much-needed experience to the students.
- Practical insights about how the industry operates and expose students to the current realities of the workplace

What is AI?

Merriam-Webster

- A branch of computer science dealing with the simulation of intelligent behavior in computers.
- The capability of a machine to imitate intelligent human behavior.

Britannica

- Artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to **reason**, **discover meaning**, **generalize**, or **learn** from past experience.

IBM Research

- By AI we mean anything that makes machines act more intelligently. Our work includes basic and applied research in **machine learning**, **deep question answering**, **search and planning**, **knowledge representation**, and **cognitive architectures**.

Four aspects of Artificially Intelligent Software

It should –

- Act humanly (The Turing Test approach)
- Think humanly (The cognitive modelling approach)
- Think rationally - Thinking with logic
- Act rationally - Always doing the right thing

The Turing Test, proposed by Alan Turing(1950), was designed to provide a satisfactory operational definition of intelligence. Turing defined intelligent behavior as the ability to achieve human-level performance in all cognitive tasks, sufficient to fool an interrogator.

Capabilities include –

Cognitive Science – the study of thought, learning, and mental organization, which draws on aspects of psychology, linguistics, philosophy, and computer modelling

Capabilities include –

- Natural language processing to enable it to communicate successfully in English (or some other human language)
- Knowledge representation to store information provided before or during the interrogation
- Automated reasoning to use the stored information to answer questions and to draw new conclusions;
- Machine learning to adapt to new circumstances and to detect and extrapolate patterns.
- Computer vision
- Robotics

Major scientific elements behind AI

- Natural Language Processing
  - *Regular expression based deterministic rules*
  - *N-gram for calculate probabilistic language model*
  - *Text classification*
  - *Information extraction (the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents)*
  - *Information retrieval (search engine e.g. google)*
- Computer Vision
- Machine Learning
  - *Prediction through regression*
  - *Classification*
- Deep learning
- Automation (RPA)
- Handling Big Data (Hadoop)
- Cloud computing (IBM Cloud)

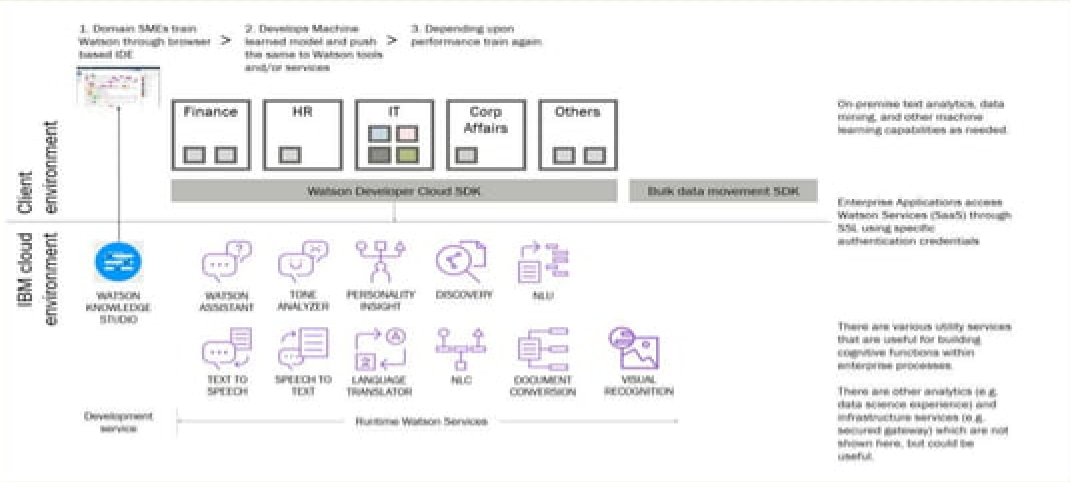
IBM Tools towards AI – APIs & Services

Category	Name	Description
AI assistant	Watson Assistant	Creates a **conversational AI solution** where user can interact with the system in 'natural' language. The **graphical tool** helps non-technical domain SMEs to contribute effectively.
Knowledge	Watson Discovery	Stores and indexes the unstructured content to support the **ranked retrieval** based on **relevancy ranking** against **query in natural language**.
	Natural Language Understanding	Applies text analytics to **extract meaning** (entities, concepts, relationship etc.) of unstructured textual content and **relate the same to enterprise ontology**.
	Discovery News	Provides **aggregated analysis** result **on news along with evidences**. Helps to stay alert for unwanted business functions.
	Knowledge Studio	Allows the domain SME's **annotate the documents** as per custom type-system (ontology), apply text analytics rules and **generate custom machine learned model** to be used by Natural Language Understanding or Watson Discovery.
Language	Language Translator	Applies **neural machine translation** to translate with ease and at a great speed. Provision are there to feed in users' own custom model.
	Natural Language Classifier	Provides provision to **train a custom model** to understand the **intent behind supplied text** and return a corresponding classification, along with a **confidence score**.
Empathy	Personality Insight	Uses **linguistic analytics** to infer individuals' personality characteristics, including Big Five, Needs, and Values, from digital communications such as email, blogs, tweets, and forum posts
	Tone Analyzer	Helps to support conversations so as to **respond to customers appropriately and at scale**. Identifies if customers are satisfied or frustrated, and if agents are polite and sympathetic.
Vision	Visual Recognition	Allows to train custom models and **identify / classify images**.
Speech	Speech to text	Converts speech to texts.
	Text to speech	Converts texts to audio output.

IBM Tools towards AI – Data Life Cycle Management

Category	Name	Description
Data (Life cycle Management)	Watson Studio	The tool to build and train AI models, and prepare and analyze data – all in one integrated environment.
	Watson Knowledge Catalog	Helps in curating structured and unstructured data. Do data profiling, classification to prepare, shape, join data as per requirement. Also, provides a collaboration platform for a team to work on the data.
	Watson Machine Learning	Build and monitor machine learning and deep learning models through easy visualization. Integrates well with Watson Studio and Watson Knowledge Catalog.

Typical architecture using Watson APIs & services



Industry usecase - Retail

	Extract	Understand	Relate	Reason	React
Customer (of a Coffee house chain) engagement experience transformation • Understand user behavior and accumulate external data • Predict user preferences • Real time offering	Cognitive analytics reads local news, pulls weather and social data and combines it with loyalty data	Data are put into context	The predictive model relates sales trend to a customer (Josie), a product (Doubleshot Iced Coffee) and to a seasonal effect (hot day).	Temporal reasoning is applied to match current weather to trend analysis to derive target customer set.	At 3 PM Josie gets an offer via SMS for a Doubleshot Iced Coffee and a free coffee for her friend
Demand Insight • Improved demand forecast • Predict inventory needs in real time	Gathers data on Social, News, Events, Weather, Economy, Competitors, Customer, supplier, Trends	Perform analytics and understand insights	The application relates various external factors (such as season data, the day of the week, important local/national events etc.) with customer behaviors, shopping habits, lifestyle.	Reasoning is applied to match the trend, seasonality and customers and come up with demand forecast.	This information may prompt a daily dashboard of suggested orders to a purchasing manager, or some AI systems may be approved to make small orders automatically without requiring human approval.

Industry usecase – Retail – contd.

	Extract	Understand	Relate	Reason	React
Personalized Product Recommendations (with help of chat service) • Personalized gifts/products recommendations • Replicate the role of a concierge	Q & A with customer in the form of chat	Understand customer need and intent before suggesting the right product.	Chat information relates customer information to the need. It then relates the need to the right product.	Reasoning is applied to understand that customers intended to go for hiking in Iceland in October and to go for commuting in Toronto in January will yield different results and intents. In the case of xyz_gift.com the reasoning is applied to tailor gift recommendation by comparing specifics provided to gifts purchased for similar recipients.	Chatbot suggests the right product suitable for the customer.

Industry usecase – Manufacturing – GE (Brilliant Factory)

General Electric is the 31st largest company in the world by revenue and one of the largest and most diverse manufacturers on the planet, making everything from large industrial equipment to home appliances. It has over 500 factories around the world and has only begun transforming them into smart facilities.

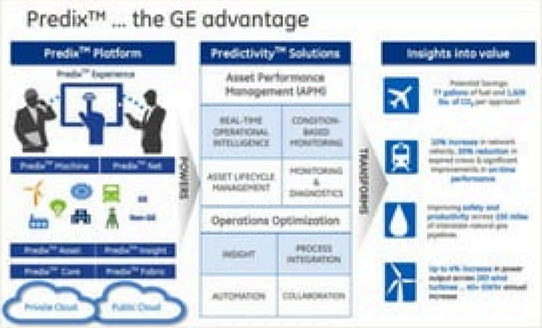
In 2015 GE launched its Brilliant Manufacturing Suite for customers, which it had been field testing in its own factories. The system takes a holistic approach of tracking and processing everything in the manufacturing process to find possible issues before they emerge and to detect inefficiencies. Their first "Brilliant Factory" was built that year in Pune, India with a $200 million investment. GE claims it improved equipment effectiveness at this facility by 18 percent.

The goal of GE's Brilliant Manufacturing Suite is to link design, engineering, manufacturing, supply chain, distribution and services into one globally scalable, intelligent system. It is powered by Predix, their industrial internet of things platform. In the manufacturing space, Predix can use sensors to automatically capture every step of the process and monitor each piece of complex equipment.

With that data, the Predix deep learning capabilities can spot potential problems and possible solutions. GE spent around $1 billion developing the system, and by 2020 GE expects Predix to process one million terabytes of data per day.

GE now has seven Brilliant Factories, powered by their Predix system, that serve as test cases. It claims positive improvements at each. For example, according to GE their system result in, their wind generator factory in Vietnam increasing productivity by 5 percent and its jet engine factory in Muskegon had a 25 percent better on-time delivery rate. They claim it has also cut unplanned downtime by 10-20 percent by equipping machines with smart sensors to detect wear.

GE (Brilliant Factory)…contd.



Industry Applications by AI elements

NLP-1.Customer service: *"How can I keep my customers happy?"*

NLP is used by computers to manipulate human language, whether to extract meaning, generate text, or for any other purpose. The interaction computer-language is categorized according to the task that needs to be accomplished: summarizing a long document, translating between two human languages, or detecting spam email are all examples of tasks that today can be decently accomplished by a machine.

While this wasn't the case 30 years ago, most of NLP today is based on machine learning i.e. statistical methods that are able to simulate what a human would do in similar circumstances.

NLP is heavily used in customer service. The interactions between customers and companies contain a lot of useful breadcrumbs that point towards the reasons for customer dissatisfaction, and the interaction itself can be cause of discontent.

In order to keep a finger on the pulse of consumers' intent, many companies now transcribe and analyze customer call recordings. They also deploy chat bots and automated online assistants to provide an immediate response to simple needs and decrease the load for customer service reps. Relevant NLP tasks include:

Speech recognition, which converts spoken language into text. Advances in deep learning over the last 10 years have allowed major players to deploy this technology in commercial systems like Siri, Google Now, Skype's translator etc. with good performances.

Question answering, which involves exactly that—answering questions posed by humans in a natural language. When in 2011 IBM's Watson outclassed the two best humans at Jeopardy!, Ken Jennings wrote on his video screen: "I, for one, welcome our new computer overlords." The task was especially hard for the machine, since the game is renowned for its convoluted and often opaque questions about general knowledge. Similar technology is used today by many companies for chatbots, both for internal (HR, operations) and external (customer service, IoT) projects.

NLP-2. Reputation monitoring: "*What are people saying about me?*"

In the 1980s, companies started using software to find patterns in their own data and make better decisions. Optimization of supply chains, inventories and warehouses, sales processes and many other applications gave rise to what we now call business intelligence. But what's inside a company's walls is not nearly as much (or as valuable) as what's outside of them.

As the cost of computation kept dropping and algorithms kept improving, businesses started adopting tools that allowed them to look beyond their databases. This kind of data is commonly referred to as external data, public data, or open-source intelligence (OSINT).

While some of this data is structured and ready to be analyzed (e.g. census data, stock prices), most of its value remains tapped in unstructured, human-generated text such as news, blog posts, forums, patents, job postings, reports, SEC filings, social media, company websites, etc. These sources contain a plethora of precious information about how competitors, customers and the market as a whole are evolving.

An example of how this kind of data can be used is reputation monitoring. It's no secret that most customers check reviews online before buying a product, whether it's a phone or a falafel. According to the most recent BrightLocal survey, 92 percent of customers read online reviews and 86 percent won't buy a product with fewer than 3 out of 5 stars.

And as consumers have started voicing their complaints on Twitter and Facebook, reputation monitoring and management has become a top priority for businesses. Companies can now scan the entire web for mentions of their brand and products and recognize cases when they should take action.

Relevant NLP tasks for this application include:

Sentiment analysis, which determines the attitude, emotional state, judgment or intent of the writer. This is done by either assigning a polarity to the text (positive, neutral or negative) or trying to recognize the underlying mood (happy, sad, calm, angry…). What about the times when multiple attitudes need to be accounted for in the same sentence? For example, "The pizza was amazing, but the waiter was awful!"

In the case above, the text is split into clauses, and polarity and mood are assessed for each. So, the previous text becomes something like:

["clause": "The pizza was amazing",
"polarity": 0.92,
"mood": "happy"],

["clause": "but the waiter was awful!",
"polarity": -0.95,
"mood": "angry"]

Coreference resolution, which connects pronouns to the right objects. This is a hard task, but it's essential to interpret the text correctly. For example, if a customer writes: "A guy from the dealer called to ask if I liked my new car. Well, no man, it sucks?", it's important to recognize that "it" refers to the car and not the guy. The customer is complaining about the product, not the service. As Elon Musk says, "Brand is just a perception". Right now, your customers are talking about you, influencing the way other consumers perceive your brand. NLP can help you stay on top of your in-the-trenches reputation.

NLP-3. Ad placement: "*Who is interested in my product?*"

Media buying is usually the largest line in a company's advertising budget, so any targeting that can be done to ensure that ads are presented to the right eyeballs is of paramount importance. While in the past marketers have focused on demographics (race, economic status, sex, age, etc.) and psychographics (values, personality, attitudes, opinions, etc.), they've quickly adapted to the new digital area.

Our emails, social media, e-commerce and browsing behaviors contain a lot of information about what we're really interested in. The huge potential of this kind of unstructured data is confirmed by the fact that 2 of the 10 largest companies today generate most of their revenue selling ads (Google and Facebook). Relevant NLP tasks for this application include:

Keyword matching, which checks whether words of interest are included in some text. This is one of the easiest tasks in NLP, and at the same time one of the most remunerative. While this first approximation is often good enough, its lack of sophistication can produce pretty inappropriate results.

Sense disambiguation, or identification of which sense of a word is used in a sentence. While the human brain is pretty good at this task, a computer won't necessarily find it easy to recognize that the term pounds in the sentence, "I gained 20 pounds since the wedding!", most likely refers to the unit of mass rather than the currency. This is still an open problem in NLP.

NLP-4. Market intelligence: "*What's happening with my competitors?*"

Markets are influenced by information exchange – between company/shareholders, government/citizens or simply individuals. Knowing the status of an industry is essential to developing an effective strategy, but the channels of content distribution today (RSS feeds, social media, emails) produce so much information that it's becoming hard to keep up.

This is especially true in financial markets, which is why hedge funds routinely use NLP to improve their models. Relevant NLP tasks for this application include:

Event extraction, which recognizes what's happening to an entity. M&As, employment changes, deal closings and everything else that happens to organizations or people can be extracted automatically. For example, "Howard Schultz Stepping Down As Starbucks CEO" can be parsed as:

{"company": "Starbucks",
"position": "CEO",
"person": "Howard Schultz",
"event": "end of employment"}

A structured database of events about companies, governments and people is an extremely powerful tool for analyzing the business ecosystem.

Sentence classification, or putting a sentence in a predefined set of buckets. This is often used as a first pass to extract relevant content from large repositories, like the SEC's EDGAR system. For example, "We expect a 15% increase in revenue next year," can be classified as a forward-looking statement.

NLP-5. Regulatory compliance: *"Is my product a liability?"*

A crucial example of compliance is pharmacovigilance i.e, the studies done after a drug has been marketed (phase IV of clinical trials) to gather information on its side effects. A lot of information about adverse drug events (ADEs) resides in what's called unstructured clinical narratives, or patients' reports about their health.

Pharma companies extract ADEs from 1) electronic health records (EHRs), social media and forums for patients complaining about a problem, and 2) web search trends and patterns for patients looking for a solution. Relevant NLP tasks for this application include:

Named entity recognition (NER), which extracts the names of drugs, diseases, patients, and pharma companies using rule-based or statistical methods. Applying NER to a sentence will be able to convert it from, "Valium makes me sleepy," to "[drug] makes me [symptom]."

Relation detection, used to identify the context in which the ADE is mentioned. This is often done with frames, or patterns of words that correspond to a concept, e.g. "I felt [symptom] after taking [drug]" is a pattern that matches the presence of side effects.

By applying these two tasks consecutively, we've identified that the patient is linking the drug Valium to the side effect sleepy. Both these tasks benefit from using ontologies i.e. structured domain knowledge (in this case biomedical) that provides the object dictionary and the relations between objects.

Computer vision – 1. Automotive

Some of the most famous applications of computer vision has been done by Tesla with their Autopilot function. The automaker launched its driver-assistance system back in 2014 with only a few features, such as lane centering and self-parking, but it's set to accomplish fully self-driving cars sometime in 2018.

Features like Tesla's Autopilot are possible thanks to startups such as Mighty AI, which offers a platform to generate accurate and diverse annotations on the datasets to train, validate, and test algorithms related to autonomous vehicles.

Computer vision – 2. Retail

Computer vision has made a splash in the retail industry as well. Amazon Go store opened up its doors to customers on January 22 this year. It's a partially automated store that has no checkout stations or cashiers. By utilizing computer vision, deep learning, and sensor fusion customers are able to simply exit the store with products of their choice and get charged for their purchases through their Amazon account. The technology is not 100% perfect yet, as several official tests of the store's technology showed that some items were left out of the final bill. However, it's an impressive step in the right direction.

A startup called Mashgin is working on a solution similar to Amazon Go. The company is working on a self-checkout kiosk that uses computer vision, 3D reconstruction, and deep learning to scan several items at the same time without the need of barcodes. The product claims to reduce check out time by up to 10x. Their main customers are cafeterias and dining halls operated by Compass Group.

Computer vision – 3. Financial Service

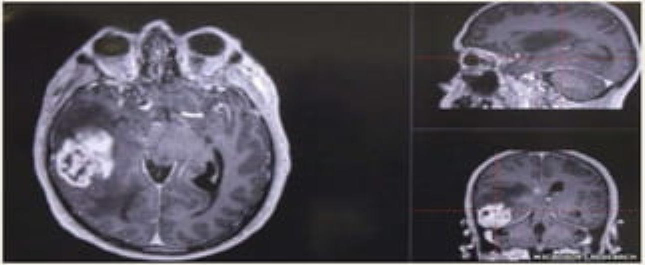
Although computer vision has not yet proved to be a disruptive technology in the world of insurance and banking, a few big players have implemented it in the onboarding of new customers. In 2016, a Spanish banking group BBVA, introduced a new way of signing up for their services. New customers could get a bank account within minutes by uploading a photo of their ID and a selfie. BBVA utilized computer vision technology to analyze the photos. Number26, an online bank based in Germany is also working on similar technology, planning to introduce it to their future clients late in 2018.

Computer vision – 4. Healthcare

In healthcare, computer vision has the potential to bring in some real value. While computers won't completely replace healthcare personnel, there is a good possibility to complement routine diagnostics that require a lot of time and expertise of human physicians but don't contribute significantly to the final diagnosis. This way computers serve as a helping tool for the healthcare personnel.

For example, Gauss Surgical is producing a real-time blood monitor that solves the problem of inaccurate blood loss measurement during injuries and surgeries. The monitor comes with a simple app that uses an algorithm that analyses pictures of surgical sponges to accurately predict how much blood was lost during a surgery. This technology can save around $10 billion in unnecessary blood transfusions every year.

One of the main challenges the healthcare system is experiencing is the amount of data that is being produced by patients. It's estimated that healthcare related data is tripled every year. Today, we as patients rely on the knowledge bank of medical personnel to analyze all that data and produce a correct diagnosis. This can be difficult at times. Microsoft's project InnerEye is working on solving parts of that problem by developing a tool that uses AI to analyze three-dimensional radiological images. The technology potentially can make the process 40 times quicker and suggest the most effective treatments.



Visual Search – Microsoft bing

In Visual search is seen as the next great search frontier, and Microsoft's Bing has tapped the power of NVIDIA GPUs to make it a reality. At the same time, they've leveraged the NVIDIA® CUDA®profiling toolchain and cuDNN to make the system more cost-effective.

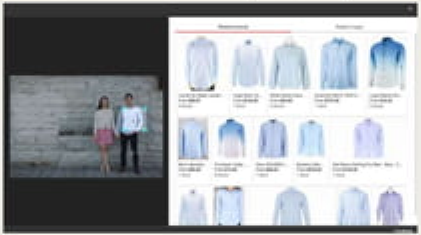
But visual search at scale is no easy matter: Instantly delivering pertinent results when users mouse over objects within photos requires massive computations by algorithms trained to classify, detect, and match the images within images.

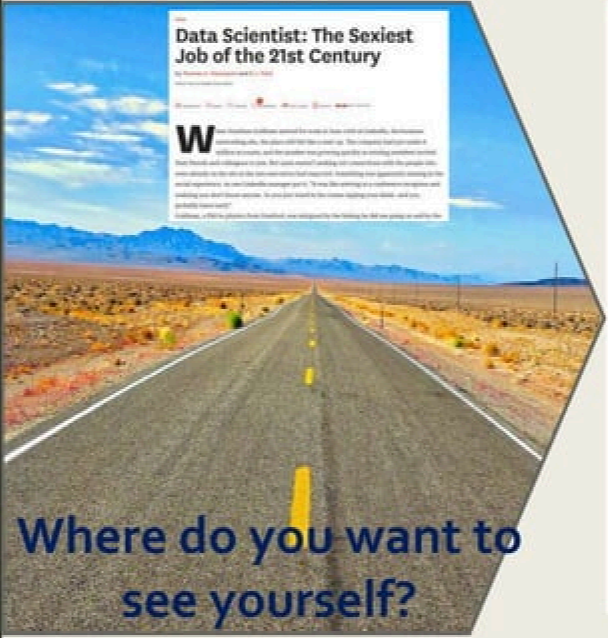
It's also well worth the effort." A picture is worth more than a thousand words," said Yan Wang, senior engineer with Bing. " When you have a picture, you're that much closer to what you're looking for."

Before now, however, it was a lengthy wait for what you were looking for. In 2015, Bing introduced image-search capabilities that enabled users to draw boxes around sub-images or click on boxes of sub-images already detected by the platform; they could then use those images as the basis of a new search.

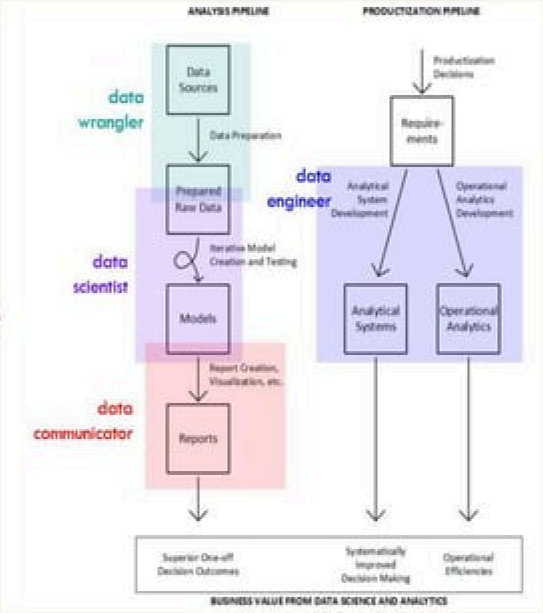
Bing sought a solution that was fast enough to keep up with user expectations. They transitioned their object-detection platform from CPUs to Azure NV-series virtual machines running NVIDIA Tesla® M60 GPU accelerators.

In doing so, Bing slashed their object-detection latency from 2.5 seconds on the CPU to 200 milliseconds. Further optimizations with NVIDIA cuDNN lowered that to 40 milliseconds, well under the threshold for an excellent user experience on most applications.





Where do you want to see yourself?



BREAK

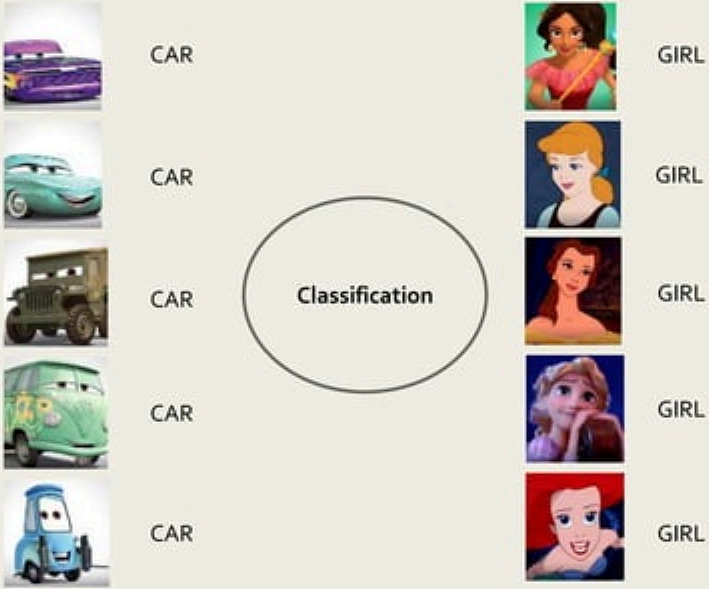
Supervised Learning

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples. In supervised learning, each example is a *pair* consisting of an input object (typically a vector) and a desired output value (also called the *supervisory signal*). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way.

Supervised learning problems can be further grouped into regression and classification problems.

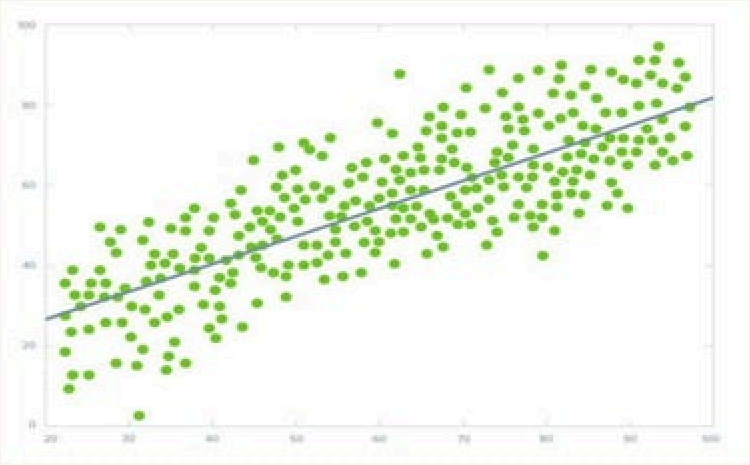
•**Classification**: A classification problem is when the output variable is a category, such as "red" or "blue" or "disease" and "no disease".

•**Regression**: A regression problem is when the output variable is a real value, such as "dollars" or "weight".


CAR
CAR
CAR
CAR
CAR
Classification
GIRL
GIRL
GIRL
GIRL
GIRL

CAR or GIRL?

Linear Regression



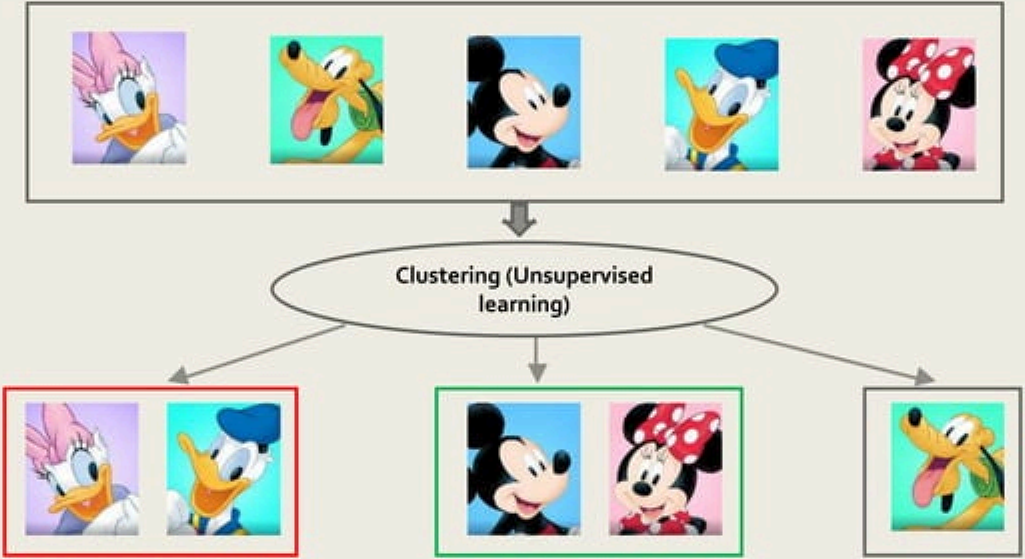
Unsupervised Learning

Unsupervised learning is a branch of machine learning that learns from test data that has not been labeled, classified or categorized. Instead of responding to feedback, unsupervised learning identifies commonalities in the data and reacts based on the presence or absence of such commonalities in each new piece of data.

Unsupervised learning problems can be further grouped into clustering and association problems.

•**Clustering**: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

•**Association**: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.


Clustering (Unsupervised learning)

Recommender System (association rule)

Some popular examples of supervised machine learning algorithm -

- Linear Regression
- Logistic Regression
- Naïve Bayes
- K- Nearest Neighbor
- Decision Tree
- Bagging (Random Forest)
- Boosting (Gradient boost, ADA boost, LGBM)
- SVM
- Neural Network

Some popular examples of unsupervised machine learning algorithm -

- Clustering (Hierarchical, K-Means)
- Dimension Reduction (PCA)
- Anomaly Detection (Fraud analysis)
- Relationship/Association Mining (Recommender system)

Linear Regression

What is Linear Regression? -

In statistics, linear regression is a linear approach to modelling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression.

What is "Linear"? –

Linearity is the property of a mathematical relationship or function which means that it can be graphically represented as a straight line. Examples are the relationship of voltage and current across a resistor (Ohm's law), or the mass and weight of an object. Proportionality implies linearity, but linearity does not imply proportionality.

What is Dependent and Independent Variable? –

In mathematical modeling, statistical modeling and experimental sciences, the values of **dependent variables** depend on the values of **independent variables**. The dependent variables represent the output or outcome whose variation is being studied. The independent variables, also known in a statistical context as **regressors**, represent inputs or causes, i.e., potential reasons for variation or, in the experimental setting, the variable controlled by the experimenter. Models and experiments test or determine the effects that the independent variables have on the dependent variables. Sometimes, independent variables may be included for other reasons, such as for their potential confounding effect, without a wish to test their effect directly.

Types of relationships

<u>Deterministic</u> <u>(or</u> **<u>functional</u>**<u>)</u> **<u>relationships –</u>**

Here is an example of a deterministic relationship. Note that the observed (*x*, *y*) data points fall directly on a line. As you may remember, the relationship between degrees Fahrenheit and degrees Celsius is known to be:

$$\text{Fahr} = \frac{9}{5}\text{Cels} + 32$$

That is, if you know the temperature in degrees Celsius, you can use this equation to determine the temperature in degrees Fahrenheit *exactly*.

Here are some examples of other deterministic relationships that students from previous semesters have shared:

- Circumference = π × diameter
- Hooke's Law: $Y = \alpha + \beta X$, where Y = amount of stretch in a spring, and X = applied weight.
- Ohm's Law: $I = V/r$, where V = voltage applied, r = resistance, and I = current.
- Boyle's Law: For a constant temperature, $P = \alpha/V$, where P = pressure, α = constant for each gas, and V = volume of gas.

For each of these deterministic relationships, the equation *exactly* describes the relationship between the two variables.

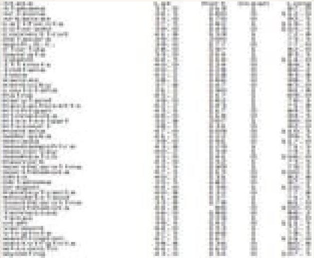
Types of relationships…contd.

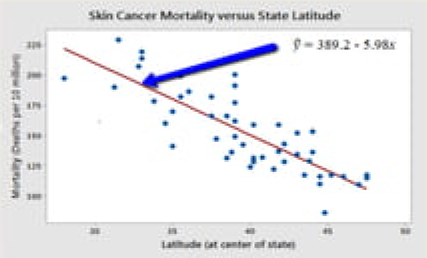
Statistical relationships –

In which the relationship between the variables is not perfect.

Here is an example of a statistical relationship.

The response variable *y* is the mortality due to skin cancer (number of deaths per 10 million people) and the predictor variable *x* is the latitude (degrees North) at the center of each of 49 states in the U.S. (The data were compiled in the 1950s, so Alaska and Hawaii were not yet states. And, Washington, D.C. is included in the data set even though it is not technically a state.)





Statistical relationships…continued… –

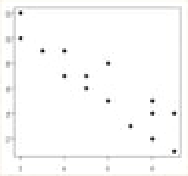
You might anticipate that if you lived in the higher latitudes of the northern U.S., the less exposed you'd be to the harmful rays of the sun, and therefore, the less risk you'd have of death due to skin cancer. The scatter plot supports such a hypothesis. There appears to be a negative linear relationship between latitude and mortality due to skin cancer, but the relationship is not perfect. Indeed, the plot exhibits some "**trend**," but it also exhibits some "**scatter**." Therefore, it is a statistical relationship, not a deterministic one.

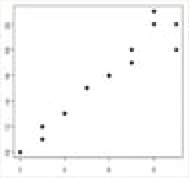
Some other examples of statistical relationships might include:

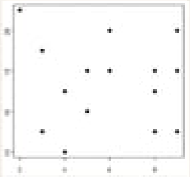
- Height and weight — as height increases, you'd expect weight to increase, but not perfectly.
- Alcohol consumed and blood alcohol content — as alcohol consumption increases, you'd expect one's blood alcohol content to increase, but not perfectly.
- Vital lung capacity and pack-years of smoking — as amount of smoking increases (as quantified by the number of pack-years of smoking), you'd expect lung function (as quantified by vital lung capacity) to decrease, but not perfectly.
- Driving speed and gas mileage — as driving speed increases, you'd expect gas mileage to decrease, but not perfectly.

Best Fitting Line

Before attempting to fit a linear model to observed data, a modeler should first determine whether or not there is a relationship between the variables of interest. This does not necessarily imply that one variable *causes* the other (for example, higher SAT scores do not *cause* higher college grades), but that there is some significant association between the two variables. A scatterplot can be a helpful tool in determining the strength of the relationship between two variables. If there appears to be no association between the proposed explanatory and dependent variables (i.e., the scatterplot does not indicate any increasing or decreasing trends), then fitting a linear regression model to the data probably will not provide a useful model.







A valuable numerical measure of association between two variables is the correlation coefficient, which is a value between -1 and 1 indicating the strength of the association of the observed data for the two variables.

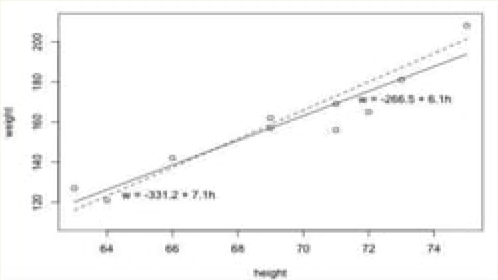
$$r = \frac{1}{n-1} \Sigma \left(\frac{x - \bar{x}}{s_x}\right)\left(\frac{y - \bar{y}}{s_y}\right)$$

A linear regression line has an equation of the form ***Y* = *a* + *bX***, where ***X*** is the explanatory variable and ***Y*** is the dependent variable. The slope of the line is ***b***, and ***a*** is the intercept (the value of ***y*** when ***x*** = 0).

Best Fitting Line…contd.

Since we are interested in summarizing the trend between two quantitative variables, the natural question arises — "what is the best fitting line?"

Looking at the plot below, which line — the solid line or the dashed line — do you think best summarizes the trend between height and weight?

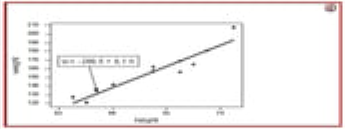


- y_i denotes the observed response for experimental unit i
- x_i denotes the predictor value for experimental unit i
- $\hat{y}_i$ is the predicted response (or fitted value) for experimental unit i

Then, the equation for the best fitting line is:

$$\hat{y}_i = b_0 + b_1 x_i$$

i	x_i	y_i	$\hat{y}_i$
1	[illegible]	[illegible]	[illegible]
2	[illegible]	[illegible]	[illegible]
3	[illegible]	[illegible]	[illegible]
4	[illegible]	[illegible]	[illegible]
5	[illegible]	[illegible]	[illegible]
6	[illegible]	[illegible]	[illegible]
7	[illegible]	[illegible]	[illegible]
8	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]
10	[illegible]	[illegible]	[illegible]



Best Fitting Line…contd.

As you can see, the size of the prediction error depends on the data point. If we didn't know the weight of student 5, the equation of the line would predict his or her weight to be -266.53 + 6.1376(69) or 157 pounds. The size of the prediction error here is 162-157, or 5 pounds.

In general, when we use $\hat{y}_i = b_0 + b_1 x_i$ to predict the actual response y_i, we make a prediction error (or residual error) of size:

$$e_i = y_i - \hat{y}_i$$

A line that fits the data "**best**" will be one for which the **n prediction errors** — one for each observed data point — **are as small as possible in some overall sense**. One way to achieve this goal is to invoke the "**least squares criterion**," which says to "minimize the sum of the squared prediction errors." That is:

- The equation of the best fitting line is: $\hat{y}_i = b_0 + b_1 x_i$
- We just need to find the values b_0 and b_1 that make the sum of the squared prediction errors the smallest it can be.
- That is, we need to find the values b_0 and b_1 that minimize:

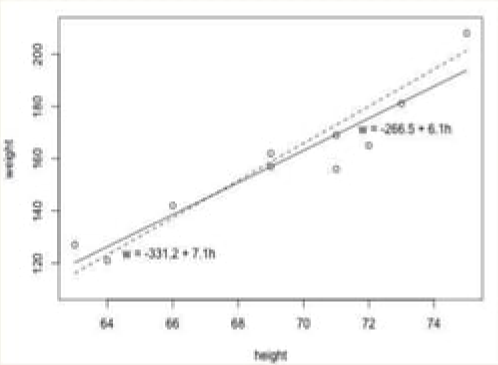
$$Q = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Here's how you might think about this quantity Q:

- The quantity $e_i = y_i - \hat{y}_i$ is the prediction error for data point i.
- The quantity $e_i^2 = (y_i - \hat{y}_i)^2$ is the squared prediction error for data point i.
- And, the symbol $\sum_{i=1}^{n}$ tells us to add up the squared prediction errors for all n data points.

Incidentally, if we didn't square the prediction error $e_i = y_i - \hat{y}_i$ to get $e_i^2 = (y_i - \hat{y}_i)^2$, the positive and negative prediction errors would cancel each other out when summed, always yielding 0.
Now, being familiar with the least squares criterion, let's take a fresh look at our plot again. In light of the least squares criterion, which line do you now think is the best fitting line?

Best Fitting Line…contd.



$w = -331.2 + 7.1\,h$ (the dashed line)					
i	x_i	y_i	$\hat{y}_i$	$(y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
1	63	127	116.1	10.9	118.81
2	64	121	123.2	-2.2	4.84
3	66	142	137.4	4.6	21.16
4	69	157	158.7	-1.7	2.89
5	69	162	158.7	3.3	10.89
6	71	156	172.9	-16.9	285.61
7	71	169	172.9	-3.9	15.21
8	72	165	180.0	-15.0	225.00
9	73	181	187.1	-6.1	37.21
10	75	208	201.3	6.7	44.89
					766.5

$w = -266.53 + 6.1376\,h$ (the solid line)					
i	x_i	y_i	$\hat{y}_i$	$(y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
1	63	127	120.139	6.8612	47.076
2	64	121	126.276	-5.2764	27.840
3	66	142	138.552	3.4484	11.891
4	69	157	156.964	0.0356	0.001
5	69	162	156.964	5.0356	25.357
6	71	156	169.240	-13.2396	175.287
7	71	169	169.240	-0.2396	0.057
8	72	165	175.377	-10.3772	107.686
9	73	181	181.515	-0.5148	0.265
10	75	208	193.790	14.2100	201.924
					597.4

Best Fitting Line...contd.

The formulas are determined using methods of calculus. We minimize the equation for the sum of the squared prediction errors:

$$Q = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2$$

(that is, take the derivative with respect to b_0 and b_1, set to 0, and solve for b_0 and b_1) and get the "**least squares estimates**" for b_0 and b_1:

$$b_0 = \bar{y} - b_1 \bar{x}$$

and:

$$b_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Because the formulas for b_0 and b_1 are derived using the least squares criterion, the resulting equation — $\hat{y}_i = b_0 + b_1 x_i$ — is often referred to as the "**least squares regression line**," or simply the "**least squares line**." It is also sometimes called the "**estimated regression equation**." Incidentally, note that in deriving the above formulas, we made no assumptions about the data other than that they follow some sort of linear trend.

We can see from these formulas that the least squares line passes through the point $(\bar{x}, \bar{y})$, since when $x = \bar{x}$, then $y = b_0 + b_1 \bar{x} = \bar{y} - b_1 \bar{x} + b_1 \bar{x} = \bar{y}$.

Simple Linear Regression Assumptions

1. The error ε is a random variable with mean of zero.
2. The variance of ε, denoted by σ^2, is the same for all values of the independent variable.
3. The values of ε are independent.
4. The error ε is a normally distributed random variable.

Error Variance

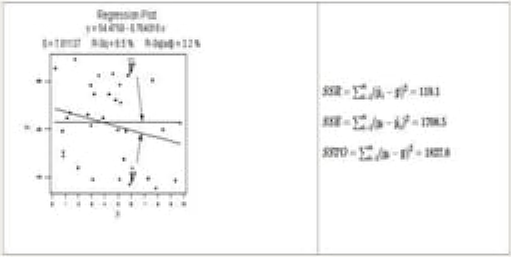
- sample variance:

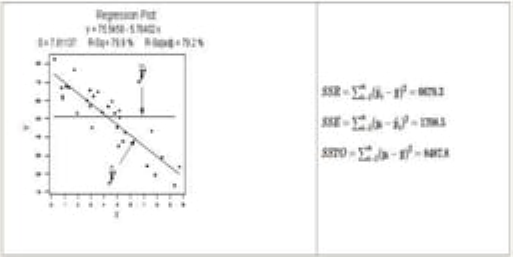
$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

- mean square error:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}$$

The Coefficient of Determination, r-squared





- SSR is the "regression sum of squares" and quantifies how far the estimated sloped regression line, $\hat{y}_i$, is from the horizontal "no relationship line," the sample mean or $\bar{y}$.
- SSE is the "error sum of squares" and quantifies how much the data points, y_i, vary around the estimated regression line, $\hat{y}_i$.
- SSTO is the "total sum of squares" and quantifies how much the data points, y_i, vary around their mean, $\bar{y}$.

$$SSTO = SSR + SSE.$$

$$r^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

(Pearson) Correlation Coefficient r

The correlation coefficient r is directly related to the coefficient of determination r^2 in the obvious way. If r^2 is represented in decimal form, *e.g.* 0.39 or 0.87, then all we have to do to obtain r is to take the square root of r^2:

$$r = \pm\sqrt{r^2}$$

One advantage of r is that it is unitless, allowing researchers to make sense of correlation coefficients calculated on different data sets with different units. The "unitless-ness" of the measure can be seen from an alternative formula for r, namely:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Another formula for r that you might see in the regression literature is one that illustrates how the correlation coefficient r is a function of the estimated slope coefficient b_1:

$$r = \frac{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \times b_1$$

In general, there is no nice practical operational interpretation for r as there is for r^2. You can only use r to make a statement about the strength of the linear relationship between x and y. In general:

- If $r = -1$, then there is a perfect negative linear relationship between x and y.
- If $r = 1$, then there is a perfect positive linear relationship between x and y.
- If $r = 0$, then there is no linear relationship between x and y.

All other values of r tell us that the relationship between x and y is not perfect. The closer r is to 0, the weaker the linear relationship. The closer r is to -1, the stronger the negative linear relationship. And, the closer r is to 1, the stronger the positive linear relationship. As is true for the r^2 value, what is deemed a large correlation coefficient r value depends greatly on the research area.

R-squared Cautions

Caution # 1

The coefficient of determination r^2 and the correlation coefficient r quantify the strength of a *linear* relationship. It is possible that r^2 = 0% and r = 0, suggesting there is no linear relation between x and y, and yet a perfect curved (or "curvilinear" relationship) exists.

Caution # 2

A large r^2 value should not be interpreted as meaning that the estimated regression line fits the data well. Another function might better describe the trend in the data.

Caution # 3

The coefficient of determination r^2 and the correlation coefficient r can both be greatly affected by just one data point (or a few data points).

Caution # 4

Correlation (or association) does not imply causation.

Caution # 5

Ecological correlations — correlations that are based on rates or averages — tend to overstate the strength of an association.

Caution # 6

A "statistically significant" r^2 value does not imply that the slope β_1 is meaningfully different from 0.

Caution # 7

A large r^2 value does not necessarily mean that a useful prediction of the response y_{new}, or estimation of the mean response μ_Y, can be made. It is still possible to get prediction intervals or confidence intervals that are too wide to be useful.

Some Maths/Stats components which are used in industry on a daily basis

Maths/Stats component	Sample usage in AI/Machine Learning
Mean, median, mode	In data cleaning and preprocessing. Helps in replacing null/NaN values. Applied for all ML algorithms.
Variance and Standard deviation	For measuring errors. e.g. applied in linear regression.
Skewness and Kurtosis	Data visualization and Data preparation
Histogram	Data visualization
Box plot	Measures of the distribution of data 1.Median of the data 2.Lower quartile 3.Upper quartile 4.Smallest observation 5.Larget observation

Some Maths/Stats components which are used in industry on a daily basis…contd.

Maths/Stats component	Sample usage in AI/Machine Learning
Confidence interval	Useful for any algorithm output. For example we can take a classification example.
Hypothesis testing (t-test and p-value)	Reviewing linear regression output.
ANOVA	Same approach is taken to measure r-square in linear regression
Normal distribution	One of the application is linear regression. Errors are assumed to be normally distributed.
Probability/Bayes theorem	One of the application is Naïve Bayes classifier

Some Maths/Stats components which are used in industry on a daily basis…contd.

Maths/Stats component	Sample usage in AI/Machine Learning
Linear Algebra	1. Dataset and Data Files 2. Images and Photographs 3. One-Hot Encoding 4. Linear Regression 5. Regularization 6. Principal Component Analysis 7. Singular-Value Decomposition 8. Latent Semantic Analysis 9. Recommender Systems 10. Deep Learning
Calculus	Partial derivatives in Gradient Descent

Classwork

student_height_weight.txt

student_height_weight.ipynb

Homework

Linear Regression Solution.ipynb

Housing Value Data

train.csv

test.csv


Thank
You
Shukria
Dhanyavadagalu
Vinaka
köszönö
Kiitos
Maake
Asante
Manana
Dankon
Matondo
감사합니다
Dankscheen
спасибо
Kam Sah Hammida
Mauruuru
Biyan
Diolch i Chi
Terima Kasih
Taiku
Tack
Dank Je
Blagodaram
Ngiyabonga
Dziekuje
Chokrane
Grazie
Arigato
Juspaxar
Mochchakkeram
Gracias
Khap
Paldies
Tingki
நன்றி
Bedankt
Ďakujem
धन्यवाद
cảm ơn bạn
Grazas
Ua Tsaug Rau Koj
Gratias Tibi
Kia Ora
Khun
Obrigado
Děkuji
Nuwun
Nirringrazzjak
Di Ou Mèsi
ありがとう
Suksama
Hvala
Rahmat
Matur
谢谢
Хвала
Welalin
Kop
Djiere Dieuf
Misaotra
Danke
Merci
Go Raibh Maith Agat
Tuke
Eskerrik Asko
Salamat
ขอขอบคุณคุณ
Najis

External references and credits -

- https://dzone.com/
- https://www.analyticsvidhya.com/
- https://newonlinecourses.science.psu.edu/statprogram/
- https://www.scipy-lectures.org/
- https://www.wikipedia.org/
- https://www.disney.com/
- Subhendu Dey (IBM India)
- https://www.techemergence.com/
- https://indatalabs.com/
- https://blogs.nvidia.com/