Date: 8/1/2022

❖ Task 1: Type of Distributions and their meaning + examples:

➢ Discrete Distributions:

- Uniform Distribution : all outcomes have same probability
- Bernoulli Distribution: events having two outcomes
- Binomial Distribution: two outcomes (like in bernoulli) but with many iterations
- Poisson's Distribution: used to test out how unusual an event frequency is for a given interval.

➢ Continuous Distributions:

- Normal Distribution
- T-Distribution: a small sample approximation of normal distribution
- Chi-squared: asymmetric, consists of only non-negative values, doesn't mirror real-life events, often used in hypothesis testing to help determine goodness of fit.
- Exponential Distribution: for events rapidly changing early on
- Logistic Distribution: useful in forecast analysis for determining a cut-off point for a successful outcome

❖ Task 2: How to change to normal distribution

Box-Cox Transformation: At the core of the Box Cox transformation is an exponent, lambda (λ), which varies from -5 to 5. All values of λ are considered and the optimal value for your data is selected; The "optimal value" is the one which results in the best approximation of a normal distribution curve. The transformation of Y has the form:

$$y(\lambda) = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, & \text{if } \lambda \neq 0; \\ \log y, & \text{if } \lambda = 0. \end{cases}$$

This test only works for positive data. However, Box and Cox did propose a second formula that can be used for negative y-values:

$$y(\boldsymbol{\lambda}) = \begin{cases} \frac{(y+\lambda_2)^{\lambda_1}-1}{\lambda_1}, & \text{if } \lambda_1 \neq 0; \\ \log(y+\lambda_2), & \text{if } \lambda_1 = 0. \end{cases}$$

❖ Task 3: Design Patterns and when they are chosen (code)

- ❖ Creational patterns: associated with control mechanisms of creating objects. The basic mode of forming an object may be problematic in some projects and may lead to unnecessary complexity in some areas. Creational patterns are supposed to prevent from occurring problems and introduce more control over creating objects. Their task is to separate the processes of creation, completion and representation of an object. There are five well-known design patterns possible to implement:

  - ➢ Abstract Factory Pattern
  - ➢ Builder Pattern
  - ➢ Factory Method Pattern
  - ➢ Prototype Pattern
  - ➢ Singleton Pattern

- ❖ Structural Patterns: the most important feature of these patterns is to facilitate the operation and design applications through finding an easy way to realize dependencies between entities. Due to these patterns, it is easier to design applications which contain independent class libraries.The following structural patterns are one of the best well-known ones:

  - ➢ Adapter Pattern
  - ➢ Decorator Pattern
  - ➢ Facade Pattern
  - ➢ Proxy Pattern
  - ➢ Composite Pattern

- ❖ Behavioural Patterns:introduce flexibility to solutions connected with inter-objects communication. They are focused on allocating specific roles and duties between objects in communication. This kind of patterns are:

  - ➢ Iterator Pattern
  - ➢ Observer Pattern
  - ➢ Command Pattern
  - ➢ Strategy Pattern
  - ➢ Template Method Pattern

❖ Task 4: Real life Example to hypothesis testing:

-Example 1: Biology

Hypothesis tests are often used in biology to determine whether some new treatment, fertilizer, pesticide, chemical, etc. causes increased growth, stamina, immunity, etc. in plants or animals.

For example, suppose a biologist believes that a certain fertilizer will cause plants to grow more during a one-month period than they normally do, which is currently 20 inches. To test this, she applies the fertilizer to each of the plants in her laboratory for one month.

She then performs a hypothesis test using the following hypotheses:

- H_0: $\mu = 20$ inches (the fertilizer will have no effect on the mean plant growth)
- H_A: $\mu > 20$ inches (the fertilizer will cause mean plant growth to increase)

If the p-value of the test is less than some significance level (e.g. $\alpha = .05$), then she can reject the null hypothesis and conclude that the fertilizer leads to increased plant growth.

-Example 2: Clinical Trials

Hypothesis tests are often used in clinical trials to determine whether some new treatment, drug, procedure, etc. causes improved outcomes in patients.

For example, suppose a doctor believes that a new drug is able to reduce blood pressure in obese patients. To test this, he may measure the blood pressure of 40 patients before and after using the new drug for one month.

He then performs a hypothesis test using the following hypotheses:

- H_0: $\mu_{after} = \mu_{before}$ (the mean blood pressure is the same before and after using the drug)
- H_A: $\mu_{after} < \mu_{before}$ (the mean blood pressure is less after using the drug)

If the p-value of the test is less than some significance level (e.g. $\alpha = .05$), then he can reject the null hypothesis and conclude that the new drug leads to reduced blood pressure.

-Example 3: Advertising Spend

Hypothesis tests are often used in business to determine whether or not some new advertising campaign, marketing technique, etc. causes increased sales.

For example, suppose a company believes that spending more money on digital advertising leads to increased sales. To test this, the company may increase money spent on digital advertising during a two-month period and collect data to see if overall sales have increased.

They may perform a hypothesis test using the following hypotheses:H0: $\mu_{after} = \mu_{before}$ (the mean sales is the same before and after spending more on advertising)

- H_A: $\mu_{after} > \mu_{before}$ (the mean sales increased after spending more on advertising)

If the p-value of the test is less than some significance level (e.g. $\alpha = .05$), then the company can reject the null hypothesis and conclude that increased digital advertising leads to increased sales.

-Example 4: Manufacturing

Hypothesis tests are also used often in manufacturing plants to determine if some new process, technique, method, etc. causes a change in the number of defective products produced.

For example, suppose a certain manufacturing plant wants to test whether or not some new method changes the number of defective widgets produced per month, which is currently 250. To test this, they may measure the mean number of defective widgets produced before and after using the new method for one month.

They can then perform a hypothesis test using the following hypotheses:

- H_0: $\mu_{after} = \mu_{before}$ (the mean number of defective widgets is the same before and after using the new method)

- H_A: $\mu_{after} \neq \mu_{before}$ (the mean number of defective widgets produced is different before and after using the new method)

If the p-value of the test is less than some significance level (e.g. $\alpha = .05$), then the plant can reject the null hypothesis and conclude that the new method leads to a change in the number of defective widgets produced per month.

❖ Task 5: How to use probability for risk taking in decision making in data science

❖ Task 6: Use of Integration in AI

Integral calculus is used in many places throughout machine learning and deep learning. Probability theory is the source of many if not most of the integrals that appear in machine learning. Probability distributions are used to represent data sources and to build statistical models. When the data is drawn probabilistically from a continuous space, then average values over the data are computed by integrals. This is called the *expectation* of a random variable, written as E[X]. $\mathbf{E[X]=\int_\Omega X(\omega)P(d\omega)}$, where **P** is a probability measure over a space **Ω**. It is a Lebesgue integral, which is equal to the Riemann integral.

The loss function being optimized are defined as expectations, and hence as integrals. The mean squared error loss results from fitting a Gaussian variable. Stochastic gradient descent estimates the gradient of an expectation (integral) over the network output. In Reinforcement Learning, various methods maximize *expected rewards*, i.e $E[R]=\int R(s,a)dP(s,a)$ for a reward R(*s*,*a*) depending on state *s* and action *a*. Policy gradients differentiate this integral, and Deep Q-Learning optimizes the integral using Bellman equations. But the bottom line is that integration is everywhere under the surface.

In some areas of ML, integration shows up in full form above the surface. The field of generative models explicitly designs networks of stochastic variables, many of them continuous. Inference within these networks requires solving or estimating a wide variety of integrals, many of them quite complex. Take a look, for instance, at any of the papers by Michael Jordan, Zoubin Ghahramani or Radford Neal, and you'll find them overflowing with integrals.

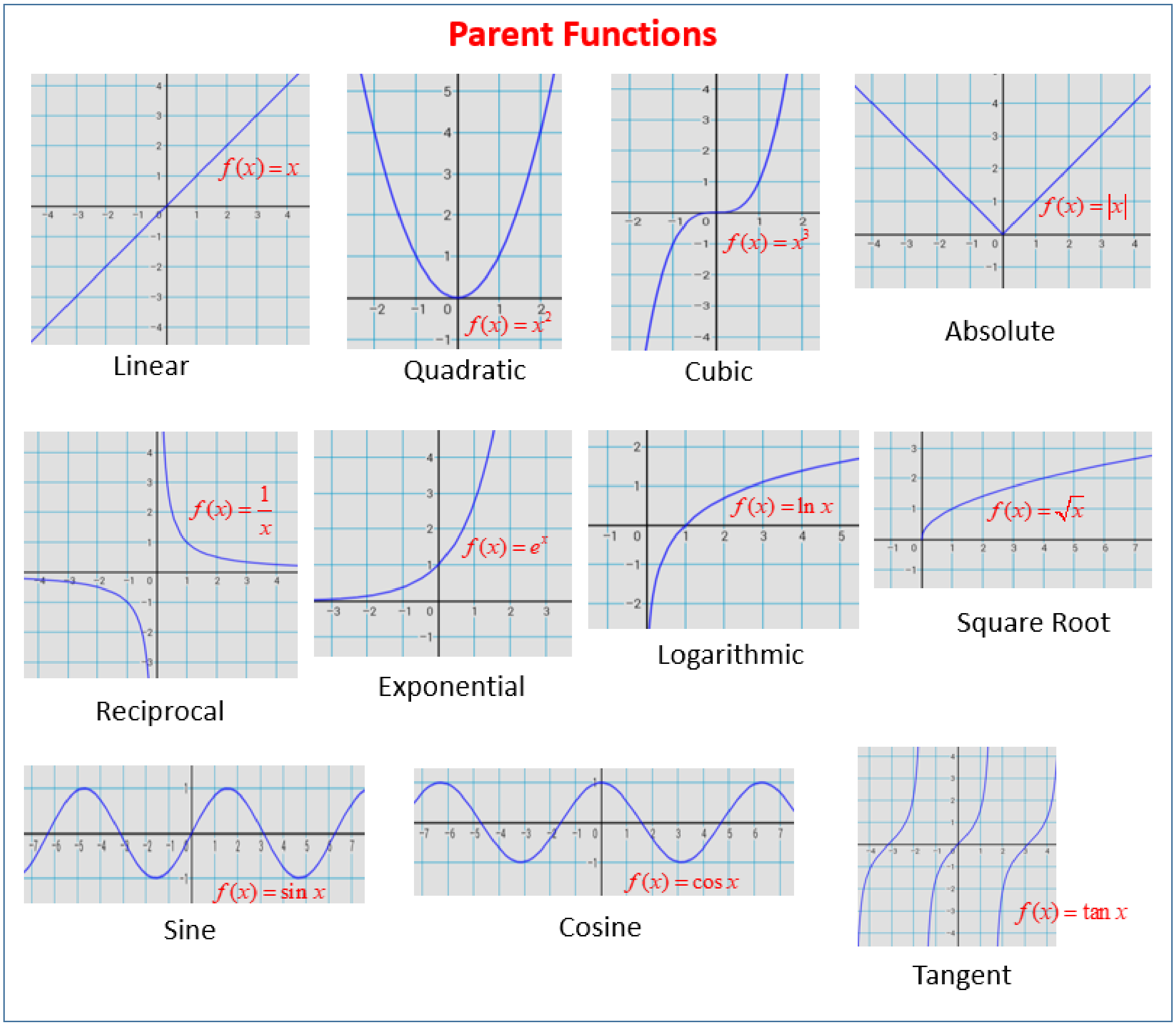
❖ Task 7: Differentiation Rules and examples:

Common Functions	Function	Derivative
Constant	c	0
Line	x	1
	ax	a
Square	x^2	2x
Square Root	√x	$(½)x^{-½}$
Exponential	e^x	e^x
	a^x	$\ln(a)\ a^x$
Logarithms	ln(x)	1/x
	$\log_a(x)$	1 / (x ln(a))
Trigonometry (x is in radians)	sin(x)	cos(x)
	cos(x)	−sin(x)
	tan(x)	$\sec^2(x)$
Inverse Trigonometry	$\sin^{-1}(x)$	$1/\sqrt{(1-x^2)}$
	$\cos^{-1}(x)$	$-1/\sqrt{(1-x^2)}$

	$\tan^{-1}(x)$	$1/(1+x^2)$

Rules	**Function**	**Derivative**
Multiplication by constant	cf	cf′
Power Rule	x^n	nx^{n-1}
Sum Rule	f + g	f′ + g′
Difference Rule	f - g	f′ − g′
Product Rule	fg	f g′ + f′ g
Quotient Rule	f/g	$\frac{f' g - g' f}{g^2}$
Reciprocal Rule	1/f	$-f'/f^2$
Chain Rule (as "Composition of Functions")	f º g	(f′ º g) × g′
Chain Rule (using ′)	f(g(x))	f′(g(x))g′(x)
Chain Rule (using d/dx)	dy/dx= dy/du x du/dx	

❖ Task 8: Equations and their graphs



❖ Task 9: How to view data with more than 3 dimensions in 2 or 3 dimensions

Dimensionality reduction is the process of reducing the number of random variables under consideration, by obtaining a set of principal variables. It can be divided into feature selection and feature extraction.

- Feature selection: In this, we try to find a subset of the original set of variables, or features, to get a smaller subset which can be used to model the problem. It usually involves three ways:
    1. Filter
    2. Wrapper
    3. Embedded
- Feature extraction: This reduces the data in a high dimensional space to a lower dimension space, i.e. a space with lesser no. of dimensions.

The various methods used for dimensionality reduction include:

- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)
- Generalized Discriminant Analysis (GDA)

❖ Task 10: Solution for unemployment due to AI

-Educational and Training Reforms:

Many high technology and skilled job positions are currently going unfilled because our twentieth century education system is not providing the dynamic and updated training necessary to fill those positions. The old model of concentrating education and training into the first quarter-century of life must shift to a life-long training paradigm in which workers are continuously being trained and updated to match new employment opportunities and technologies. For example, the ACT Foundation has envisioned a future life-long learning system called the Learning Ledger, which keeps an account of each individual's educational and learning credits achieved throughout their lives, with educational units available from a wide variety of institutions and organizations.

-Tax Policy and Financial Incentives:

Tax policy and other government financial incentives could be used to promote job creation. For example, a tax credit could be provided for each new job provided, creating an incentive for an employer to favor a human over machine worker in marginal cases. Tax policy incentives could also be targeted at employees rather than just employers, such as by for example expanding the earned income tax credit (EITC) to provide greater rewards for individuals on government support to add to their income by taking on new jobs.

-Technological Innovation:

Even though technological innovation (particularly in the computer, robot and AI fields) is the driver of technological unemployment, many types of technological innovation are continuing to generate new jobs. For example, personalized medicine is greatly increasing demand for genetic counselors, drones are creating thousands of new jobs for designers and operators, 3D printers are creating brand new markets for CAD files, and artificial intelligence is creating strong demand for coders and software engineers. While some have suggested that the threat of technological unemployment might justify slowing technological innovation, such an approach would not only deny society the benefits of new technologies, but would also likely do more harm than good with respect to employment opportunities. We need smarter innovation, that in part is directed at creating valuable new job and career opportunities.

❖ Task 11: How to know the rate of change of a rate of change

❖ Task 12: Cases where there are more than one rate of change

- ❖ **Task 13: Performance difference between statistical and non statistical models with an example**