جامعة المستقبل
AL MUSTAQBAL UNIVERSITY

قسم الانظمة الطبية الذكية

المرحلة الثالثة

Subject: Artificial Intelligence AII

Class: Third

Lecturers: Dr. Muneera Abed Hadi, M. Sc. Rouaa Safi , M. Sc. Ansam Ali

University of Information Technology
and Communications
College of Medical Informatics
Intelligent Medical Systems Department

Machine Learning Course

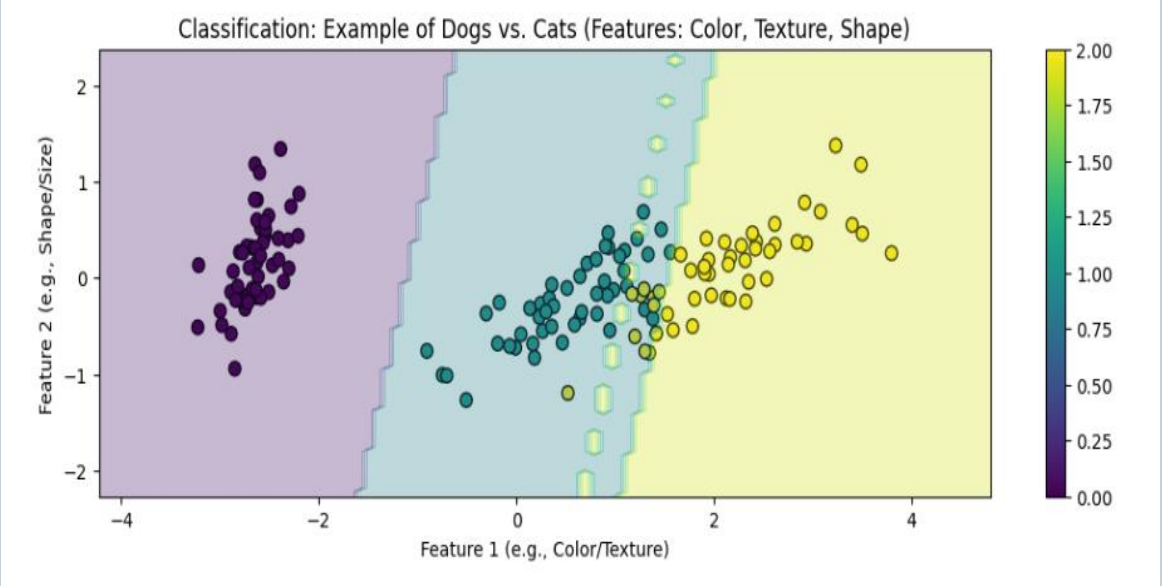
Lecture Five/P1

By: Dr. Muneera Abed Hmdi

Getting Started with Classification

- ***Classification*** teaches a machine to sort things into categories. It learns by looking at examples with labels (like emails marked “spam” or “not spam”). After learning, it can decide which category new items belong to, like identifying if a new email is ***spam or not***. For example, a classification model might be trained on dataset of images labeled as either dogs or cats and it can be used to predict the class of new and unseen images as dogs or cats based on their features such as color, texture and shape.

- ***Explaining classification in ml, horizontal axis represents the combined values of color and texture features. Vertical axis represents the combined values of shape and size features.***



Getting Started with Classification

- Each colored dot in the plot represents an ***individual image***, with the color indicating whether the model predicts the image to be a ***dog or a cat.***

- The ***shaded areas*** in the plot show the ***decision boundary***, which is the ***line or region*** that the model uses to decide which category (dog or cat) an image belongs to. The model classifies images on one side of the boundary as dogs and on the other side as cats, based on their features.

Types of Classification

- When we talk about classification in machine learning, we're talking about the process of sorting data into categories based on specific features or characteristics. There are different types of classification problems depending on how many categories (or classes) we are working with and how they are organized. There are three main classification types in machine learning:

1. Binary Classification

- The goal in binary classification is to sort the data into two distinct categories. An example, a system that sorts emails into either spam or not spam. It works by looking at different features of the email like certain keywords or sender details and decides whether it's spam or not. It only chooses between these two options.

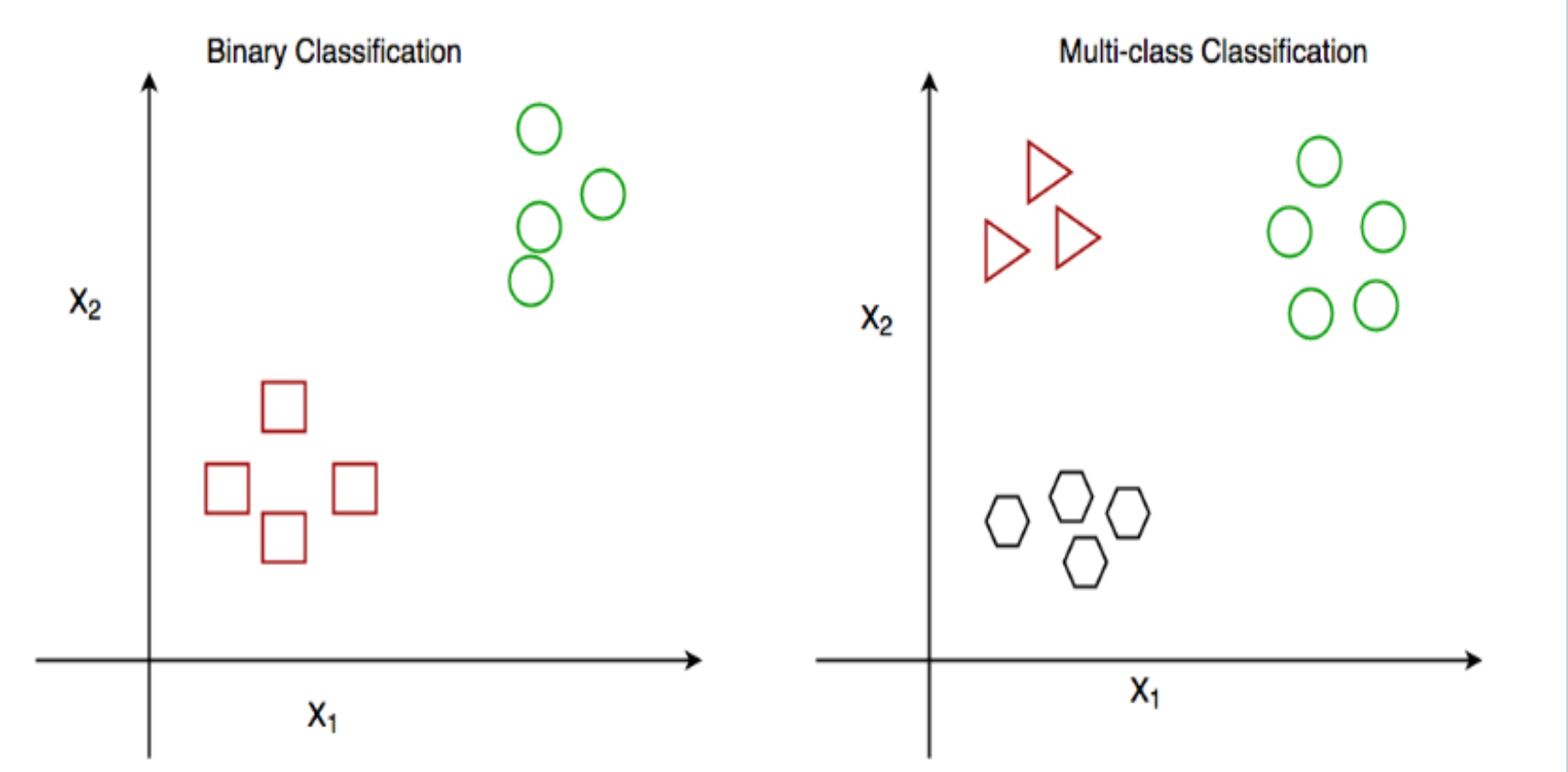
Types of Classification

2. Multiclass Classification

- In this type, instead of just two categories, the data needs to be sorted into ***more than two categories***. The model picks the one that best matches the input. Think of an image recognition system that sorts pictures of animals into categories like ***cat, dog, and bird***.

- Basically, machine looks at the features in the image (like ***shape, color, or texture***) and chooses which animal the picture is most likely to be based on the training it received.

Types of Classification



Types of Classification

3. Multi-Label Classification:

- In multi-label classification, single piece of data can belong to multiple categories at once. Unlike multiclass classification where each data point belongs to only one class, multi-label classification allows datapoints to belong to multiple classes. An example, a movie recommendation system could tag a movie as both action and comedy. The system checks various features (like ***movie plot, actors, or genre tags***) and assigns multiple labels to a single piece of data, rather than just one.

  - ***Multilabel classification is relevant in specific use cases, but not as crucial for a starting overview of classification.***

How does Classification in Machine Learning Work?

- In machine learning, classification works by training a model to ***learn patterns from labeled data***, so it can predict the category or class of new, unseen data. How it works?

- ***Data Collection:*** You start with a dataset where each item is labeled with the correct class (for example, “cat” or “dog”).
- ***Feature Extraction:*** The system identifies features (like color, shape, or texture) that help distinguish one class from another. These features are what the model uses to make predictions.
- ***Model Training:*** Classification – machine learning algorithm uses the labeled data to learn how to map the features to the correct class. It looks for patterns and relationships in the data.

How does Classification in Machine Learning Work?

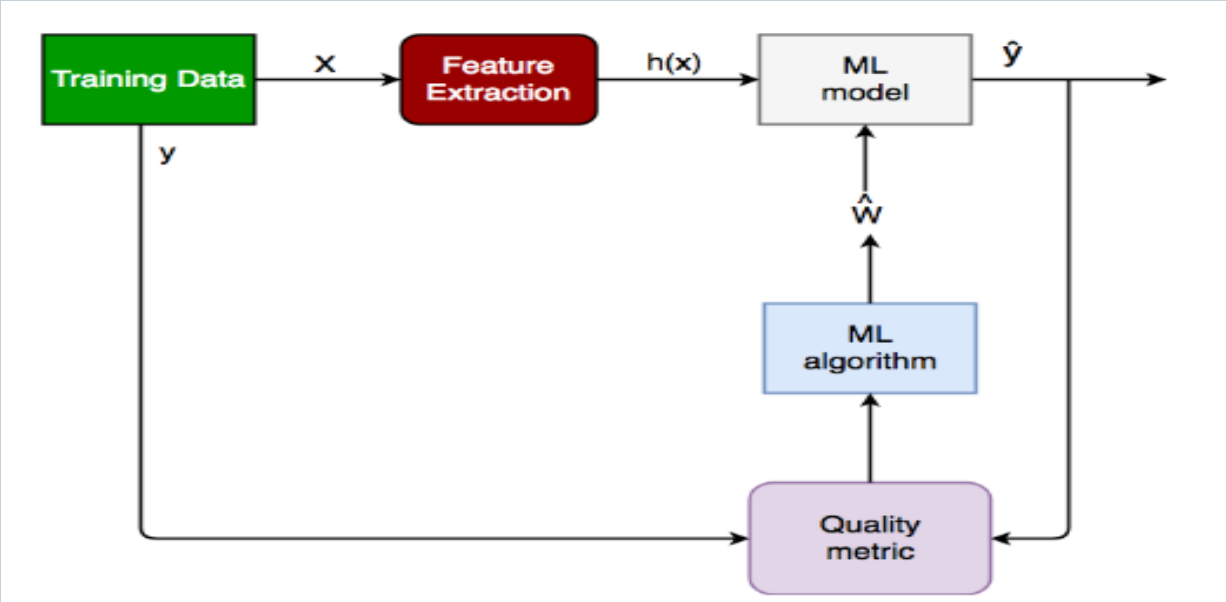
- In machine learning, classification works by training a model to ***learn patterns from labeled data***, so it can predict the category or class of new, unseen data. How it works?

- ***Model Evaluation:*** Once the model is trained, it's tested on new, unseen data to check how accurately it can classify the items.

- ***Prediction:*** After being trained and evaluated, the model can be used to predict the class of new data based on the features it has learned.

- ***Model Evaluation:*** Evaluating a classification model is a key step in machine learning. It helps us check how well the model performs and how good it is at handling new, unseen data. Depending on the problem and needs we can use different metrics to measure its performance.

Classification Machine Learning



- If the quality metric is ***not satisfactory***, the ML algorithm or ***hyperparameters can be adjusted***, and the model is ***retrained***. This iterative process continues until a ***satisfactory performance is achieved***. In short, classification in machine learning is all about using existing labeled data to teach the model ***how to predict the class of new, unlabeled data based on the patterns it has learned***.

key Characteristics of Classification Models

- ***Class Separation:*** Classification relies on distinguishing between distinct classes. The goal is to learn a model that can separate or categorize data points into predefined classes based on their features.

- ***Decision Boundaries:*** The model draws decision boundaries in the feature space to differentiate between classes. These boundaries can be linear or non-linear.

- ***Sensitivity to Data Quality:*** Classification models are sensitive to the quality and quantity of the training data. Well-labeled, representative data ensures better performance, while noisy or biased data can lead to poor predictions.

- ***Handling Imbalanced Data:*** Classification problems may face challenges when one class is underrepresented. Special techniques like resampling or weighting are used to handle class imbalances.

- ***Interpretability:*** Some classification algorithms, such as Decision Trees, offer higher interpretability, meaning it's easier to understand why a model made a particular prediction.

key Characteristics of Classification Models

- ***There are various types of classifiers algorithms such as:***

1. Linear Classifiers: Linear classifier models create a linear decision boundary between classes. They are simple and computationally efficient. Some of the linear classification models are as follows:

- ***Logistic Regression***
- ***Support Vector Machines having kernel = 'linear'***
- ***Single-layer Perceptron***
- ***Stochastic Gradient Descent (SGD) Classifier***

key Characteristics of Classification Models

- ***There are various types of classifiers algorithms such as:***

2. Non-linear Classifiers: Non-linear models create a non-linear decision boundary between classes. They can capture more complex relationships between input features and target variable. Some of the non-linear classification models are as follows:

- ***K-Nearest Neighbors***
- ***Kernel SVM***
- ***Naive Bayes***
- ***Decision Tree Classification***
- ***Random Forests,***
- ***Multi-layer Artificial Neural Networks***

Examples of Machine Learning Classification in Real Life

- ***Classification algorithms are widely used in many real-world applications across various domains, including:***
  - ***Email spam filtering***.
  - ***Credit risk assessment:*** Algorithms predict whether a loan applicant is likely to default by analyzing factors such as credit score, income, and loan history. This helps banks make informed lending decisions and minimize financial risk.
  - ***Medical diagnosis :*** Machine learning models classify whether a patient has a certain condition (e.g., cancer or diabetes) based on medical data such as test results, symptoms, and patient history. This aids doctors in making quicker, more accurate diagnoses, improving patient care.

Examples of Machine Learning Classification in Real Life

- ***Classification algorithms are widely used in many real-world applications across various domains, including:***
  - ***Image classification:*** Applied in fields such as facial recognition, autonomous driving, and medical imaging.
  - ***Sentiment analysis:*** Determining whether the sentiment of a piece of text is positive, negative, or neutral. Businesses use this to understand customer opinions, helping to improve products and services.
  - ***Fraud detection:*** Algorithms detect fraudulent activities by analyzing transaction patterns and identifying anomalies crucial in protecting against credit card fraud and other financial crimes.
  - ***Recommendation systems:*** Used to recommend products or content based on past user behavior, such as suggesting movies on Netflix or products on Amazon. This personalization boosts user satisfaction and sales for businesses.

Regression in machine learning

- ***Regression*** in machine learning refers to a supervised learning technique where the goal is to predict a continuous numerical value based on one or more independent features. It finds relationships between variables so that predictions can be made. we have two types of variables present in regression:

  - ***Dependent Variable (Target): The variable we are trying to predict e.g. house price.***

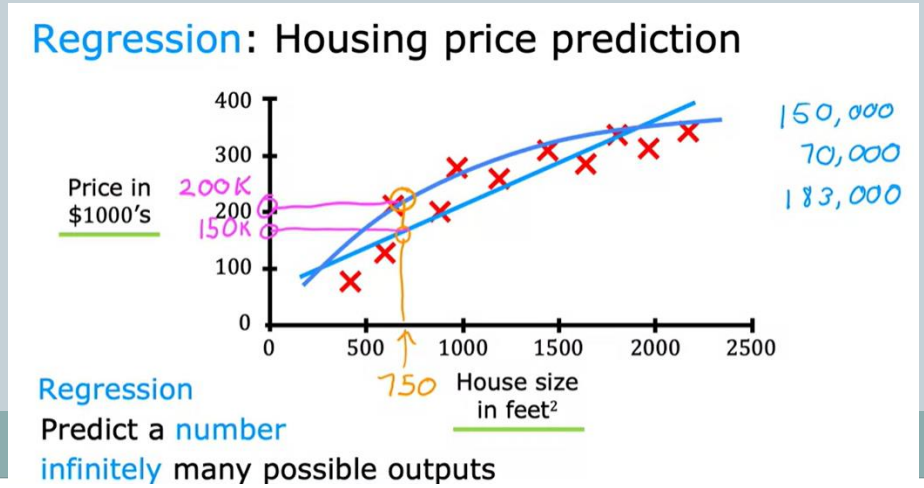
  - ***Independent Variables (Features): The input variables that influence the prediction e.g. locality, number of rooms.***

Types of Regression

- ***Regression*** can be classified into different types based on the number of predictor variables and the nature of the relationship between variables:

1. Simple Linear Regression

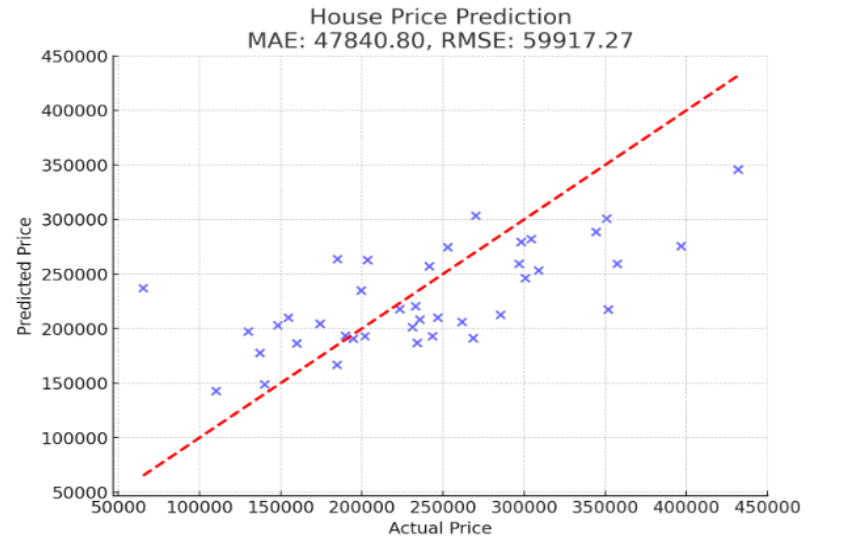
- Linear regression is one of the simplest and most widely used statistical models. This assumes that there is a linear relationship between the independent and dependent variables. This means that the change in the dependent variable is proportional to the change in the independent variables. For example, predicting the price of a house based on its size.

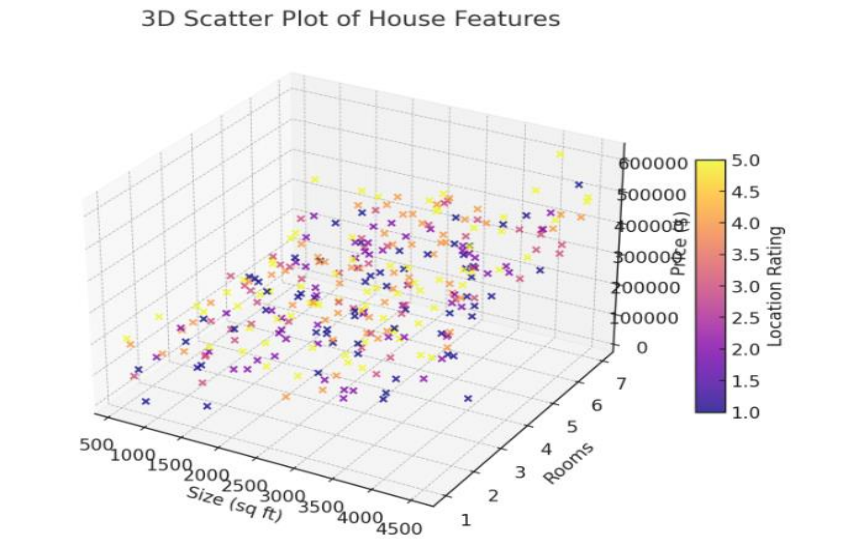


Types of Regression

2. Multiple Linear Regression

- Multiple linear regression extends simple linear regression by using multiple independent variables to predict target variable. For example, predicting the price of a house based on multiple features such as size, location, number of rooms, etc.





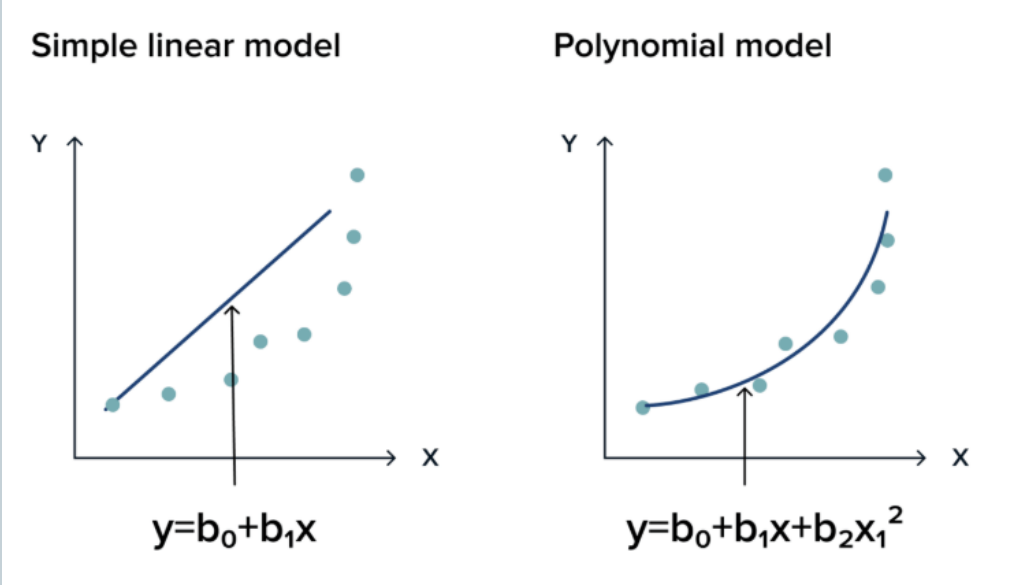
- **a scatter plot comparing actual vs. predicted house prices**
- **It shows the relationship between size, number of rooms, and price, with color representing location rating**

Types of Regression

3. Polynomial Regression

- ***Polynomial regression*** is used to model with non-linear relationships between the dependent variable and the independent variables. It adds polynomial terms to the linear regression model to capture more complex relationships. For example, when we want to predict a non-linear trend like population growth over time we use polynomial regression.

- In the equation, ***Y*** is the dependent variable, ***X*** is the independent variable, and b_0–b_n are the parameters you can optimize.



Applications of Regression

- ***Predicting prices:*** Used to predict the price of a house based on its size, location and other features.
- ***Forecasting trends:*** Model to forecast the sales of a product based on historical sales data.
- ***Identifying risk factors:*** Used to identify risk factors for heart patient based on patient medical data.
- ***Making decisions:*** It could be used to recommend which stock to buy based on market data.

Advantages vs. Disadvantages of Regression

- ***Advantages of Regression***
  - ***Easy to understand and interpret.***
  - ***Robust to outliers.***
  - ***Can handle both linear relationships easily.***

- ***Disadvantages of Regression***
  - ***Assumes linearity.***
  - ***Sensitive to situation where two or more independent variables are highly correlated with each other i.e. multicollinearity.***
  - ***May not be suitable for highly complex relationships.***