- **Machine Learning With Python**
- **Welcome**
- Video
- In this course you'll learn how Machine Learning is used in many key fields and industries

- **Introduction to Machine Learning**

- **Python for Machine Learning**
- **Supervised Vs Unsupervised**

- **Quiz : Intro to Machine Learning**
- Supervised learning deals with unlabeled data, while unsupervised learning deals with labelled data.
- False
- The “Regression” technique in Machine Learning is a group of algorithms that are used for

- When comparing Supervised with Unsupervised learning, is this sentence True or False?
- In contrast to Supervised learning, Unsupervised learning has more models and more evaluation methods that can be used in order to ensure the outcome of the model is accurate
- False

- **Linear Regression**
- Introduction to Regression

- **Simple Linear Regression**
- In this video, we'll be covering linear regression. You don't need to know any linear algebra to understand topics in linear regression. This high-level introduction will give you enough background information on linear regression to be able to use it effectively on your own problems. So let's get started
-
-
-
-
-
- Model evaluation in Regression Models

- Evaluation Metrics in Regression Models

- **Lab : Simple Linear Regression**
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%202/ML0101EN-Reg-Simple-Linear-Regression-Co2.ipynb
- ✅
-
-
-
- **Multiple linear regression**
-
- **Lab : Multiple Linear Regression**

- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%202/ML0101EN-Reg-Mulitple-Linear-Regression-Co2.ipynb
- **Non Linear Regression**
- Video
- **Lab Polynomial Regression**
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%202/ML0101EN-Reg-Polynomial-Regression-Co2.ipynb
- **Lab Non Linear Regression**
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%202/ML0101EN-Reg-NoneLinearRegression.ipynb
- *Let's learn about non linear regressions and apply an example on python. In this notebook, we fit a non-linear model to the data points corresponding to China's GDP from 1960 to 2014.*
- **Quiz : Regression**
- **Multiple Linear Regression is appropriate for:**
- Predicting tomorrow's rainfall amount based on the wind speed and temperature
- **Which of the following is the meaning of "Out of Sample Accuracy" in the context of evaluation of models?**
- "Out of Sample Accuracy" is the percentage of correct predictions that the model makes on data that the model has NOT been trained on
- **When should we use Multiple Linear Regression?**
- When there are multiple dependent variables
- When we would like to predict impacts of changes in independent variables on a dependent variable.
- **Which of the following statements are TRUE about Polynomial Regression?**
- Polynomial regression models can fit using the Least Squares method.
- Polynomial regression fits a curve line to your data.
- Polynomial regression can use the same mechanism as Multiple Linear Regression to find the parameters.
- **Which sentence is NOT TRUE about Non-linear Regression?**
- Nonlinear regression is a method to model non linear relationship between the dependent variable and a set of independent variables.
- For a model to be considered non-linear, y must be a non-linear function of the parameters.
- **Non-linear regression must have more than one dependent variable**
- **WEEK 3**
- **K Nearest Neighbours**
- **Introduction to Classification**
-
- **K Nearest Neighbours**
-
- **Evaluation metrics in classification**

- **Lab : KNN**
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%203/ML0101EN-Clas-K-Nearest-neighbors-CustCat.ipynb
- **Decision Trees**
- **Introduction to Decision Trees**
- **Building Decision Trees**
- **Lab : Decision Trees**
- In this lab exercise, you will learn a popular machine learning algorithm, Decision Tree. You will use this classification algorithm to build a model from historical data of patients, and their respond to different medications. Then you use the trained decision tree to predict the class of a unknown patient, or to find a proper drug for a new patient. Click HERE to download the lab notebook (.ipynb)
- **Logistic Regression**
- Intro to Logistic Regression

In this video, we'll learn a machine learning method called Logistic Regression which is used for classification. In examining this method, we'll specifically answer these three questions. What is logistic regression? What kind of problems can be solved by logistic regression? In which situations do we use logistic regression?

- Logistic Regression Vs Linear Regression

we will learn the difference between linear regression and logistic regression. We go over linear regression and see why it cannot be used properly for some binary classification problems. We also look at the sigmoid function, which is the main part of logistic regression.

- Logistic Regression Training

we will learn more about training a logistic regression model. Also, we will be discussing how to change the parameters of the model to better estimate the outcome. Finally, we talk about the cost function and gradient descent in logistic regression as a way to optimize the model

- **Lab : Logistic Regression**
- In this notebook, you will learn Logistic Regression, and then, you'll create a model with telecommunications data to predict when its customers will leave for a competitor, so that you can take some action to retain the customer. Click HERE to download the lab notebook (.ipynb)
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%203/ML0101EN-Clas-Logistic-Reg-churn.ipynb
- **Support Vector Machine**
- Video

Hello and welcome. In this video, we will learn a machine learning method called, Support

Vector Machine, or SVM, which is used for classification.

In this notebook, you will use SVM (Support Vector Machines) to build and train a model using human cell records, and classify cells to whether the samples are benign or malignant. Click HERE to download the lab notebook (.ipynb)

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%203/ML0101EN-Clas-SVM-cancer.ipynb

Week 3 : Quiz

Classification

Question 1: Which one IS NOT a sample of classification problem?

- To predict the category to which a customer belongs to.
- To predict whether a customer switches to another provider/brand.
- **To predict the amount of money a customer will spend in one year.**
- To predict whether a customer responds to a particular advertising campaign or not.

Question 2: Which of the following statements are TRUE about Logistic Regression? (select all that apply)

- **Logistic regression can be used both for binary classification and multi-class classification**
- **Logistic regression is analogous to linear regression but takes a categorical/discrete target field instead of a numeric one.**
- **In logistic regression, the dependent variable is binary.**

Which of the following examples is/are a sample application of Logistic Regression? (select all that apply)

- **The probability that a person has a heart attack within a specified time period using person's age and sex.**
- **Customer's propensity to purchase a product or halt a subscription in marketing applications.**
- **Likelihood of a homeowner defaulting on a mortgage.**

- Estimating the blood pressure of a patient based on her symptoms and biographical data.

Which one is TRUE about the kNN algorithm?

- kNN is a classification algorithm that takes a bunch of unlabelled points and uses them to learn how to label other points.
- **kNN algorithm can be used to estimate values for a continuous target.**

What is "information gain" in decision trees?

- It is the information that can decrease the level of certainty after splitting in each node.
- **It is the entropy of a tree before split minus weighted entropy after split by an attribute.**
- It is the amount of information disorder, or the amount of randomness in each node.

End of week 3

- **Week 4 : Clustering**
- **k-Means Clustering**
-
- **Intro to Clustering**

In this video we'll give you a high level introduction to clustering, its applications, and different types of clustering algorithms

- **Intro to k means**
- **More on K means**

In this video, we'll look at k-Means accuracy and characteristics.

- **Lab : k means**

Despite its simplicity, the K-means is vastly used for clustering in many data science applications, especially useful if you need to quickly discover insights from unlabeled data. In this notebook, you learn how to use k-Means for customer segmentation. Click HERE to download the lab notebook (.ipynb)

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%204/ML0101EN-Clus-K-Means-Customer-Seg.ipynb

Hierarchical clustering

Intro to Hierarchical clustering

More on hierarchical clustering

Hello and welcome. In this video, we'll be covering more details about hierarchical clustering. Let's get started. Let's look at agglomerative algorithm for hierarchical clustering.

lets compare hierarchical clustering with K-means. K-means is more efficient for large data sets. In contrast to K-means, hierarchical clustering does not require the number of cluster to be specified. Hierarchical clustering gives more than one partitioning depending on the resolution or as K-means gives only one partitioning of the data. Hierarchical clustering always generates the same clusters, in contrast with K-means, that returns different clusters each time it is run, due to random initialization of centroids.

Lab : Agglomerative Clustering

In this lab, we will be looking at Agglomerative clustering, which is more popular than Divisive clustering. We will also be using Complete Linkage as the Linkage Criteria. Click HERE to download the lab notebook (.ipynb)

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%204/ML0101EN-Clus-Hierarchical-Cars.ipynb

- **Density Based Clustering**
- DBSCAN
- 7 min

Hello and welcome. In this video, we'll be covering DB scan. A density-based clustering algorithm which is appropriate to use when examining spatial data.

- **Lab : DBSCAN Clustering**

Density-based Clustering locates regions of high density that are separated from one another by regions of low density. Density, in this context, is defined as the number of points within a specified radius.

In this section, the main focus will be manipulating the data and properties of DBSCAN and observing the resulting clustering. Click HERE to

download the lab notebook (.ipynb)

- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%204/ML0101EN-Clus-DBSCN-weather.ipynb

- Quiz : Clustering

Machine Learning with Python Coursera Quiz Answers Week 4

Question 1: Which statement is NOT TRUE about k-means clustering?

- k-means divides the data into non-overlapping clusters without any cluster-internal structure.
- The objective of k-means, is to form clusters in such a way that similar samples go into a cluster, and dissimilar samples fall into different clusters.
- **As k-means is an iterative algorithm, it guarantees that it will always converge to the global optimum.**

Question 2: Which of the following are characteristics of DBSCAN? Select all that apply.

- **DBSCAN can find arbitrarily shaped clusters.**
- **DBSCAN can find a cluster completely surrounded by a different cluster.**
- **DBSCANhas a notion of noise, and is robust to outliers.**
- **DBSCAN does not require one to specify the number of clusters such as k in k-means**

Question 3: Which of the following is an application of clustering?

- Customer churn prediction
- Price estimation
- **Customer segmentation**
- Sales prediction

Question 4: Which approach can be used to calculate dissimilarity of objects in clustering?

- Minkowski distance
- Euclidian distance
- Cosine similarity
- **All of the above**

Question 5: How is a center point (centroid) picked for each cluster in k-means?

- **We can randomly choose some observations out of the data set and use these observations as the initial means.**
- **We can create some random points as centroids of the clusters.**

- We can select it through correlation analysis.
- **End of week 4**
- **Week 5**
-
- **Content Based Recommendation Engines**
- **Intro to Recommender systems**
-
- **Content based recommender systems**
- Lab : Content based Recommendation systems

Recommendation systems are a collection of algorithms used to recommend items to users based on information taken from the user. These systems have become ubiquitous can be commonly seen in online stores, movies databases and job finders. In this notebook, we will explore Content-based recommendation systems and implement a simple version of one using Python and the Pandas library. Click HERE to download the lab notebook (.ipynb)

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%205/ML0101EN-RecSys-Content-Based-movies.ipynb

- **Collaborative filtering**
- Video
- In this video, we'll be covering a recommender system technique called collaborative filtering

Lab: Collaborative Filtering on Movies

Recommendation systems are a collection of algorithms used to recommend items to users based on information taken from the user. These systems have become ubiquitous can be commonly seen in online stores, movies databases and job finders. In this notebook, we will explore recommendation systems based on Collaborative Filtering and implement simple version of one using Python and the Pandas library. Click HERE to download the lab notebook (.ipynb)

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%205/ML0101EN-

RecSys-Collaborative-Filtering-movies.ipynb

Quiz : Recommender System

Question 1: What is/are the advantage/s of Recommender Systems ?

- Recommender Systems provide a better experience for the users by giving them a broader exposure to many different products they might be interested in.
- Recommender Systems encourage users towards continual usage or purchase of their product
- Recommender Systems benefit the service provider by increasing potential revenue and better security for its consumers.

What is a content-based recommendation system?

- **Content-based recommendation system tries to recommend items to the users based on their profile built upon their preferences and taste.**

What is the meaning of "Cold start" in collaborative filtering?

The difficulty in recommendation when we have new user, and we cannot make a profile for him, or when we have a new item, which has not got any rating yet.

What is a "Memory-based" recommender system?

In memory based approach, we use the entire user-item dataset to generate a recommendation system.

What is the shortcoming of content-based recommender systems?

- **Users will only get recommendations related to their preferences in their profile, and recommender engine may never recommend any item with other characteristics.**

End of Week 5

Week 6

Final Project

Reading : How to do final project? PDF

Reading: Instructions for Final Peer Graded Assignment PDF

This final project will be graded by your peers who are completing this course during the same session. This project is worth 25 marks of

your total grade, broken down as follows:

1. Building model using KNN, finding the best k and accuracy evaluation (7 marks)
2. Building model using Decision Tree and find the accuracy evaluation (6 marks)
3. Building model using SVM and find the accuracy evaluation (6 marks)
4. Building model using Logistic Regression and find the accuracy evaluation (6 marks)

Peer-graded Assignment: The best classifier

Rishab Rawat(Best Classifier).ipynb

Final Exam

What is the subfield of computer science that gives "computers the ability to learn without being explicitly programmed."?

Machine Learning

Regression/Estimation, Classification, Clustering, and Associations are all examples of what

ML Techniques

Which type of regression model can by transformed into a linear regression model using the Least Squares method?

Least Squares method

Which one IS a sample of classification problem?

To predict whether a customer responds to a particular advertising campaign or not.

To predict the category to which a customer belongs to.

To predict whether a customer switches to another provider/brand.

~~To predict the amount of money a customer will spend in one year.~~

Which of the following statements are TRUE about Logistic Regression? (select all that apply)

1 / 1 point

Logistic regression can be used both for binary classification and multi-class classification

Logistic regression is analogous to linear regression but takes a categorical/discrete target field instead of a numeric one.

In logistic regression, the dependent variable is binary.

TRUE about k-means clustering?

The objective of k-means, is to form clusters in such a way that similar samples go into a cluster, and dissimilar samples fall into different clusters.

k-means divides the data into non-overlapping clusters without any cluster-internal structure.

~~As k-means is an iterative algorithm, it guarantees that it will always converge to the global optimum.~~

characteristics of DBSCAN

DBSCAN can find arbitrarily shaped clusters.

DBSCAN can find a cluster completely surrounded by a different cluster.

DBSCAN has a notion of noise, and is robust to outliers.

DBSCAN does not require one to specify the number of clusters such as k in k-means

A Reccomender system provides a better experience for the user by giving them a broader exposure to many different products they might be interested in.

Question 17) What is a content-based recommendation system?

- **Content-based recommendation system tries to recommend items to the users based on their profile built upon their preferences and taste.**

When we should use Multiple Linear Regression?

- **When we would like to identify the strength of the effect that the independent variables have on a dependent variable.**
-