

Introduction to Machine Learning using Python

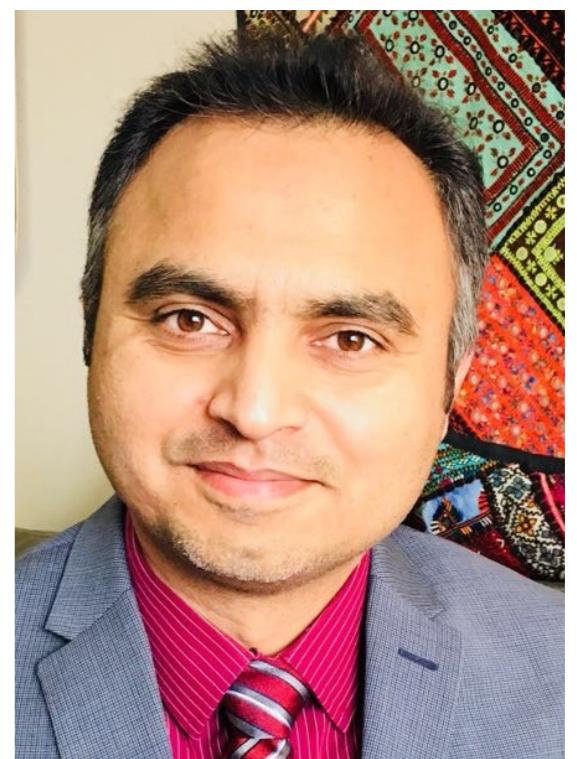
Hi guys

Welcome to the Machine Learning Section!

We have set the foundations and covered Python Essentials along with its powerful capabilities in Data Analysis and Data Visualization.

It's time to move on to the next extremely important section,
“Machine Learning”.

Dr. Junaid S. Qazi, PhD



Material and optional reading

In this section, we will often reference to the following resources (available for free):

1. [Machine Learning - A Probabilistic Perspective](#) by Kevin P. Murphy
2. [An Introduction to Statistical Learning](#) by Gareth James
3. [Scikit-learn official documentation](#)

Although, the lecture notes and exercises are self sufficient, However, you can always explore the suggested material to learn more.

Machine Learning: What and why?

- We are in the era of big data
- There are about 1 trillion of web pages and more are adding daily
- Around one hour of video is uploaded to YouTube every second
- Only Walmart handles more than 1M transactions per hour
- databases are continuing to grow in petabytes and more.....

Machine Learning: What and why?

- We are in the era of big data
- There are about 1 trillion of web pages and more are adding daily
- Around one hour of video is uploaded to YouTube every second
- Only Walmart handles more than 1M transactions per hour
- databases are continuing to grow in petabytes and more.....

To handle this humongous amount of data and related issues, something intelligent is required, a set of methods, that can automatically detect patterns in the data and predict future data or perform other kind of decisions.

Machine Learning provides this all!

Machine Learning: Possible applications

The applications of Machine Learning are not limited and include:

- Targeted marketing
- Real-time ads on the webpage
- Fraud detection
- Credit scoring, insurance and next best offers
- Pricing models for new products
- Text analysis
- Email spam
- Financial modelling and Business Intelligence
- Prediction of equipment failures
- Smart homes and so on.....

We will explore lots of these with hands-on projects in this course!

Machine Learning Process:

- **Data Acquiring** - Data may have lots of features / dimensions or variables, may need some cleaning (we have seen such examples in the data analysis and visualization sections)
- **Data Cleaning** - The process is separating / extracting the data that we need, re-organizing data into a right format. If needed, creating new features and discard some existing ones.
- **Splitting data into Test and Train Sets**
- **Model Training** and building on Train dataset then **Model Testing** on Test dataset - iterative process until we get the best fit to our data - The Model is Ready!
- **Model Deployment!**

Machine Learning Process:

- **Data Acquiring** - Data may have lots of features / dimensions or variables, may need some cleaning (we have seen such examples in the data analysis and visualization sections)
- **Data Cleaning** - The process is separating / extracting the data that we need, re-organizing data into a right format. If needed, creating new features and discard some existing ones.
 - Splitting data into Test and Train Sets
 - Model Training and building on Train dataset then Model Testing on Test dataset - iterative process until we get the best fit to our data - The Model is Ready!
 - Model Deployment!

Machine Learning Process:

- **Data Acquiring** - Data may have lots of features / dimensions or variables, may need some cleaning (we have seen such examples in the data analysis and visualization sections)
- **Data Cleaning** - The process is separating / extracting the data that we need, re-organizing data into a right format. If needed, creating new features and discard some existing ones.
- **Splitting data into Test and Train Sets**
- Model Training and building on Train dataset then Model Testing on Test dataset - iterative process until we get the best fit to our data - The Model is Ready!
- Model Deployment!

Machine Learning Process:

- **Data Acquiring** - Data may have lots of features / dimensions or variables, may need some cleaning (we have seen such examples in the data analysis and visualization sections)
- **Data Cleaning** - The process is separating / extracting the data that we need, re-organizing data into a right format. If needed, creating new features and discard some existing ones.
- **Splitting data into Test and Train Sets**
- **Model Training** and building on Train dataset then **Model Testing** on Test dataset - **iterative process** until we get the best fit to our data - The Model is Ready!
- **Model Deployment!**

Types of Machine Learning:

Machine Learning has mainly three types:

1. **Supervised** or Predictive Learning
2. **Unsupervised** or Descriptive Learning
3. **Reinforcement** Learning

Let's discuss little more on these algorithms one-by-one

Types of Machine Learning:

1. Supervised Learning:

In this type of learning algorithms:

- *A set of labeled data, based on known features / dimensions / variables is provided.*
- *Algorithms are trained using labeled examples.*

Simply, the learning algorithm receives a set of inputs along with the corresponding correct outputs, it then learns by comparing its actual output with correct outputs to find errors. The algorithm then modifies the model accordingly and get the best deployable model based on the training data.

Types of Machine Learning:

1. Supervised Learning:

In this type of learning algorithms:

- *A set of labeled data, based on known features / dimensions / variables is provided.*
- *Algorithms are trained using labeled examples.*

Simply, the learning algorithm receives a set of inputs along with the corresponding correct outputs, it then learns by comparing its actual output with correct outputs to find errors. The algorithm then modifies the model accordingly and get the best deployable model based on the training data.

Types of Machine Learning:

1. Supervised Learning:

Supervised learning is commonly used in application where historical data predicts likely future events.

For example, in a very simple setting, the algorithm is given height and weight as features along with labels as Male and Female. A model trained on the data with known features and labels, can be used to predict the gender of a human if the unlabeled data (features only) are provided.

Another very common example is the prediction of house price (label) using a model trained on features such as no. of rooms, area, floor and price as label.

Supervised learning problems are grouped into Classification and Regression

Types of Machine Learning:

1. Supervised Learning:

Supervised learning is commonly used in application where historical data predicts likely future events.

For example, in a very simple setting, the algorithm is given height and weight as features along with labels as Male and Female. A model trained on the data with known features and labels, can be used to predict the gender of a human if the unlabeled data (features only) are provided.

Another very common example is the prediction of house price (label) using a model trained on features such as no. of rooms, area, floor and price as label.

Supervised learning problems are grouped into Classification and Regression

Types of Machine Learning:

2. Unsupervised Learning:

- Only inputs are given with no historical labels or target values and the goal is to find the “**interesting patterns**”. This is sometime called **knowledge discovery**.
- This is much less well-defined problem, since the system is not told the right answer and the algorithm must figure out what is being shown.
- Popular techniques include self-organizing maps, nearest-neighbour mapping, k-mean clustering and singular value decomposition.

Unsupervised learning problems are largely grouped into Clustering
(discovering the inherent grouping in the data e.g. grouping customers by purchasing price) and
Association *(discovering a rule that describe a large portions of the data e.g. people who watched X also tend to watch Y)*

Types of Machine Learning:

2. Unsupervised Learning:

- Only inputs are given with no historical labels or target values and the goal is to find the “**interesting patterns**”. This is sometime called **knowledge discovery**.
- This is much less well-defined problem, since the system is not told the right answer and the algorithm must figure out what is being shown.
- Popular techniques include self-organizing maps, nearest-neighbour mapping, k-mean clustering and singular value decomposition.

Unsupervised learning problems are largely grouped into Clustering
(discovering the inherent grouping in the data e.g. grouping customers by purchasing price) and
Association *(discovering a rule that describe a large portions of the data e.g. people who watched X*
also tend to watch Y)

Types of Machine Learning:

2. Unsupervised Learning:

- Only inputs are given with no historical labels or target values and the goal is to find the “**interesting patterns**”. This is sometime called **knowledge discovery**.
- This is much less well-defined problem, since the system is not told the right answer and the algorithm must figure out what is being shown.
- Popular techniques include self-organizing maps, nearest-neighbour mapping, k-mean clustering and singular value decomposition.

Unsupervised learning problems are largely grouped into Clustering
(discovering the inherent grouping in the data e.g. grouping customers by purchasing price) and
Association *(discovering a rule that describe a large portions of the data e.g. people who watched X*
also tend to watch Y)

Types of Machine Learning:

2. Unsupervised Learning:

- Only inputs are given with no historical labels or target values and the goal is to find the “**interesting patterns**”. This is sometime called **knowledge discovery**.
- This is much less well-defined problem, since the system is not told the right answer and the algorithm must figure out what is being shown.
- Popular techniques include self-organizing maps, nearest-neighbour mapping, k-mean clustering and singular value decomposition.

Unsupervised learning problems are largely grouped into Clustering (discovering the inherent grouping in the data e.g. grouping customers by purchasing price) and Association (discovering a rule that describe a large portions of the data e.g. people who watched X also tend to watch Y)

Types of Machine Learning:

3. Reinforcement Learning:

- **Reinforcement Learning** is somewhat less commonly used and inspired by behaviourist psychology.
- Often used for robotics, gaming, operational research and navigation.
- The algorithm discovers through trial and error which actions yield the greatest reward.
- **Three primary components:** the **agent** (learner or decision maker), the **environment** (everything the agent interacts with) and **actions** (what the agent can do)
- The typical framing of Reinforcement Learning scenario is, an agent takes actions in an environment, which is interpreted into a reward and a representation of the state, which are fed back into the agent.

(The agent will reach the goal much faster by following a good policy, hence the goal of RL is to learn the best policy – consider how a baby learns to walk)

Types of Machine Learning:

3. Reinforcement Learning:

- Reinforcement Learning is somewhat less commonly used and inspired by behaviourist psychology.
- Often used for robotics, gaming, operational research and navigation.
- The algorithm discovers through trial and error which actions yield the greatest reward.
- Three primary components: the **agent** (learner or decision maker), the **environment** (everything the agent interacts with) and **actions** (what the agent can do)
- The typical framing of Reinforcement Learning scenario is, an agent takes actions in an environment, which is interpreted into a reward and a representation of the state, which are fed back into the agent.

(The agent will reach the goal much faster by following a good policy, hence the goal of RL is to learn the best policy – consider how a baby learns to walk)

Types of Machine Learning:

3. Reinforcement Learning:

- Reinforcement Learning is somewhat less commonly used and inspired by behaviourist psychology.
- Often used for robotics, gaming, operational research and navigation.
- The algorithm discovers through trial and error which actions yield the greatest reward.
- **Three primary components:** the **agent** (learner or decision maker), the **environment** (everything the agent interacts with) and **actions** (what the agent can do)
- The typical framing of Reinforcement Learning scenario is, an agent takes actions in an environment, which is interpreted into a reward and a representation of the state, which are fed back into the agent.

(The agent will reach the goal much faster by following a good policy, hence the goal of RL is to learn the best policy – consider how a baby learns to walk)

Types of Machine Learning:

3. Reinforcement Learning:

- Reinforcement Learning is somewhat less commonly used and inspired by behaviourist psychology.
- Often used for robotics, gaming, operational research and navigation.
- The algorithm discovers through trial and error which actions yield the greatest reward.
- **Three primary components:** the **agent** (learner or decision maker), the **environment** (everything the agent interacts with) and **actions** (what the agent can do)
- The typical framing of Reinforcement Learning scenario is, an agent takes actions in an environment, which is interpreted into a reward and a representation of the state, which are fed back into the agent.

(The agent will reach the goal much faster by following a good policy, hence the goal of RL is to learn the best policy – consider how a baby learns to walk)

Learning strategy in this section

We will learn range of algorithm using real data in this Machine Learning section. Our learning strategy will be practical, simple and very effective:

- **First:** Overview on the topic and optional reading
- **Second:** Demonstration lecture using Python on the topic/algorithm with data
- **Third:** Project / assignments for practice, at the end of each topic/ algorithm and then overview of solutions for the project.

Learning strategy in this section

We will learn range of algorithm using real data in this Machine Learning section. Our learning strategy will be practical, simple and very effective:

- **First:** Overview on the topic and optional reading
- **Second:** Demonstration lecture using Python on the topic/algorithm with data
- **Third:** Project / assignments for practice, at the end of each topic/ algorithm and then overview of solutions for the project.

Learning strategy in this section

We will learn range of algorithm using real data in this Machine Learning section. Our learning strategy will be practical, simple and very effective:

- **First:** Overview on the topic and optional reading
- **Second:** Demonstration lecture using Python on the topic/algorithm with data
- **Third:** Project / assignments for practice, at the end of each topic/ algorithm and then overview of solutions for the project.

Our practical teaching methodology will make the things much easier to understand and follow the steps involved. We will actually perform machine learning tasks on the datasets in this section.

Important Note

It takes time to learn, specially machine learning. You need to be patient and committed to learn and understand. Remember, Machine Learning is one of the major skill you need to become a successful data scientist.

The resource links provided in this section are great to learn the basics in the field. However, not one single resource can be considered as a reference for all machine learning topics. The lectures are prepared using all the references and much more.

You are always welcome to post your questions in Q&A forum, and we will be pleased to guide though the topic.

Important Note

It takes time to learn, specially machine learning. You need to be patient and committed to learn and understand. Remember, Machine Learning is one of the major skill you need to become a successful data scientist.

The resource links provided in this section are great to learn the basics in the field. However, not one single resource can be considered as a reference for all machine learning topics. The lectures are prepared using all the references and much more.

You are always welcome to post your questions in Q&A forum, and we will be pleased to guide though the topic.

Important Note

It takes time to learn, specially machine learning. You need to be patient and committed to learn and understand. Remember, Machine Learning is one of the major skill you need to become a successful data scientist.

The resource links provided in this section are great to learn the basics in the field. However, not one single resource can be considered as a reference for all machine learning topics. The lectures are prepared using all the references and much more.

You are always welcome to post your questions in Q&A forum, and we will be pleased to guide though the topic.

Yes, learning is fun!

Bear with me, you will find Machine Learning very interesting, fulfilling and rewarding.

Scikit Learn

[**Scikit-learn**](#) is the most popular machine learning package for Python and it has a lot of built-in algorithms. You need to install this package because we will use this package in this section.

- **conda install scikit-learn**

If you don't have Anaconda distribution installed, you can use pip .

- **pip install scikit-learn**

Once the package is installed, we are ready to move on. Let's jump on the next lecture to get hands-on in Machine Learning!

[**Scikit-learn algorithm official cheat-sheet**](#)