



SUMMER ANALYTICS

MODULE 05: DEEPER INTO MACHINE LEARNING

Hello everyone!

Kudos to you on completing the first four modules!

In this module, we will continue our journey to implement various Machine Learning models and understand their usage. Also, we will look at various evaluation metrics and their significance. This is followed by various feature extraction and transformation techniques.

Also, some optional content is added in case you need better clarity on some topics.

TOPICS:

1. Machine Learning models:

Understanding the intuition behind various Machine Learning models and their code will be essential when we try to solve a real problem.

TASK 1	SVM: Intuition	A short video to build an intuition about SVM
TASK 2	SVM: Code	Understanding SVM with the code for reference
TASK 3	Decision Trees: Concept Video Decision Trees: Concept Blog	A short blog and a video to introduce Decision Trees.
TASK 4	Decision Trees: Code	Understanding Decision Trees with the code for reference
TASK 5	Random Forest: Concept	A short video to understand the Random Forest model
TASK 6	Random Forest: Code	A code-based example of the Random Forest implementation
TASK 7	ML-Models :: Self-check	This tutorial notebook will give you hands-on experience in implementing all these models. It will also introduce you to GridSearchCV.
TASK 8	Hyperparameter Tuning	A short blog on GridSearchCV and RandomizedSearchCV

2. Evaluation Metrics:

Now that you can build ML models, you need some way to determine which model would work best for your problem. So, let's understand the various performance evaluation metrics for ML models.

TASK 1	Confusion Matrix	Basic concepts include Gradient descent and Cost Function for a Linear Regression model
TASK 2	Sensitivity, Specificity	Linear Algebra theory required for implementation
TASK 3	AUC - ROC	A summary of the process for review
TASK 4	Summary of the above three	A more generalized implementation of Linear Regression
TASK 5	Evaluation Metrics: Classification	A blog about choosing the relevant evaluation metric for Classification
TASK 6	Evaluation Metrics: Regression	A blog about choosing the relevant evaluation metric for Regression
TASK 7	Regression metrics	Another blog

3. Data Transformation and Feature Selection

Note that the domain of data transformation and feature selection is vast. It is difficult to cover it all in a small bunch of videos/blogs. This topic is very relevant and widely practiced by Data Science professionals. I'd recommend you explore it by going beyond the following material.

TASK 1	Dealing with Imbalanced Data: Blog1 Dealing with Imbalanced Data - Blog2	Data Imbalance is a common problem while training models for Classification. Here you will learn how to handle such cases.
TASK 2	Handling Skewed Data	Generally, your predictions are more reliable if the data is normally distributed (Remember Gauss?) Here you'll get an overview of what a transformation looks like.
TASK 3	Feature Selection: Blog1	When there is a huge number of feature columns in the data, some kind of Feature Selection helps in improving the model by preventing overfitting.
TASK 4	Feature Selection with ScikitLearn: Blog2	This is an easier implementation of the same. Try it yourself on the model in the self-check assignment.

3. Other useful techniques:

TASK 1	Polynomial Regression: Blog	Another type of Regression.
TASK 2	Polynomial Regression: Video	For those who prefer videos over blogs.
TASK 3	Binning	Another useful concept.
TASK 4	Binning: Code	Implementation of Binning in a real example to see how useful it is.

Congratulations, you've completed all of the learning content in Module-5!

QUIZ

This component is graded. Feel free to use the internet to understand the concepts better. Do these questions independently for the best results.

Click [here](#) to attempt the test.