



TIPS FOR WEEK 2

Find answers to these simple questions (should not take more than a few hours).

The research that you do here (in finding answers to these questions) would help you while you are working on any Machine Learning or Deep Learning problem.

Label Encoding & One Hot Encoding

1. What is Label Encoding & One-Hot Encoding? What is the need of each?
2. What happens if we do not One-Hot encode and rather only use Label Encoding while working on Algorithms like Logistic Regression or Neural Networks?
3. Is one-hot encoding required for Tree based methods?

Evaluation Metric - Classification

1. What is Precision? What is Recall?
2. What is True Positive, True Negative, False Positive & False Negative?
3. What is Confusion Matrix? Why is it important to look at Confusion matrix to infer results rather than just plain vanilla accuracy?

Normalization

1. What is Normalization and why is it required to scale features within a certain range?
Hint: <https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization-standardization/>
2. Does normalization or standardization help if we use Tree based Algorithms?
3. Do we need to normalize our target variables?

Hyperparameter Tuning

1. What are hyperparameters? Give examples!
2. What are some mechanisms for automated hyperparameter tuning in scikit-learn?
3. When do you use RandomizedSearchCV v/s GridSearchCV?
4. Are these automated methods scalable when you work on large datasets?

Simple Imputation

1. How do you handle missing, Nan or Null values in your data using Pandas?
Hint: <https://www.quora.com/How-do-I-handle-NaN-values-in-a-Pandas-Dataframe>