To

IITD-AIA Foundation of Smart Manufacturing

Date:16-07-2023

Subject: ***Weekly Progress Report for Week-8.***

Dear Sir,

Following is the required progress report of this week dated from 24-07-2023 to 30-07-2023.

Weekly Progress:

July 24 & July 25 :

Topics covered:

- The model development cycle goes through various stages, starting from data collection to model building.
- But, before exploring the data to understand relationships (in variables), It's always recommended to perform hypothesis generation.
- In machine learning, the term model accuracy refers to the measurements made to decide whether or not a certain model is the best to describe the relationship between the different problem variables.
- We often use training data (sample data) to train a model for new, unused data.
- If our model has good accuracy, it will perform well on both the training data and the new one.
- Data augmentation is a technique used to artificially increase the size of a dataset by generating new data from the existing data.

July 26:

Topics covered:

- Ensemble methods are machine learning techniques that combine the predictions of multiple individual models to produce a more accurate and robust final prediction.
- These methods can improve the overall performance of a model by leveraging the strengths of different models and mitigating their weaknesses.
- Ensemble methods are particularly useful when dealing with complex or noisy datasets, as they help reduce overfitting and improve generalization.
- There are three main types of ensemble methods:

1. Bagging: Bagging stands for Bootstrap Aggregating. It involves training multiple instances of the same model on different subsets of the training data, which are obtained by random sampling with replacement.
2. Boosting: Boosting is an iterative process that sequentially builds multiple models, each correcting the errors of its predecessor.
3. Stacking: Stacking combines multiple different models (also called base or level-0 models) by training a meta-model (level-1 model) that learns from the predictions of the base models.

July 27:

Topics covered:

- Cross-validation is a resampling technique used to evaluate the performance of a machine learning model and estimate its generalization ability on unseen data.
- It involves partitioning the available data into multiple subsets or folds, where each fold is used as both a training set and a validation set.
- This process helps to assess the model's stability and robustness by averaging the performance over multiple iterations.
- The most commonly used cross-validation method is k-fold cross-validation.

```
Mean MSE: 3.9063015139066692
Standard Deviation of MSE: 4.34553226772623 45
```

- The main advantage of cross-validation is that it provides a more accurate and unbiased estimate of the model's generalization performance compared to a single train-test split.
- It also allows you to make the most of your data, especially when the dataset is limited.
- The evaluation metric used is mean squared error (MSE).
- The code calculates the mean and standard deviation of the cross-validation scores, which provide an estimate of the model's performance and its variability.

July 28:

Topics covered:

- Regularization is a technique used in machine learning to prevent overfitting and improve the generalization of a model to new, unseen data.
- Overfitting occurs when a model performs well on the training data but fails to generalize to new data, leading to poor performance on the test or validation set.
- Regularization helps to address this issue by adding a penalty term to the loss function during model training, which discourages the model from becoming too complex.
- There are two common types of regularization used in machine learning:
    - L1 Regularization (Lasso): L1 regularization adds the absolute value of the model's weights as a penalty term to the loss function. This encourages the model to set some of the less important features' weights to exactly zero,

effectively performing feature selection. It leads to sparse weight vectors and helps to create simpler models.
    - L2 Regularization (Ridge): L2 regularization adds the squared magnitude of the model's weights as a penalty term to the loss function. Unlike L1 regularization, L2 regularization does not force any weights to be exactly zero, but it penalizes large weights, making them smaller. This results in a more balanced and smoother model.
- Regularization is usually controlled by a hyperparameter called the regularization strength (λ or alpha).
- A higher regularization strength will result in more aggressive regularization, leading to simpler models with smaller weights. Conversely, a lower regularization strength will allow the model to have larger weights, potentially leading to more complex models and overfitting.

July 29 & July 30:

Topics covered:

- Transfer learning is a machine learning technique that leverages knowledge learned from one task to improve the performance of a related but different task.
- The main idea behind transfer learning is that the knowledge gained from learning patterns and representations in the source domain can be useful for learning similar patterns in the target domain.
- By using a pre-trained model, you can benefit from the general features learned from vast amounts of data, even when you have limited data for the specific target task.
- In transfer learning, a model that has been pre-trained on a large and general dataset (source domain) is used as a starting point for a new task with a smaller dataset (target domain).
- The benefits of transfer learning include faster training times, better generalization to new data, and improved performance on the target task, especially when you have limited data.
- It is particularly valuable in scenarios where collecting large amounts of data for the target task is expensive or time-consuming.
- By adapting and fine-tuning the pre-trained model on our lathe machine dataset, you may be able to achieve better performance and quicker convergence during training.