“Cross Validation” Module Transcript

Chapter 1

Intro, Topics Covered, & Learning Outcomes

Howdy, folks! My name is Austin Hambrick, and I'm a machine learning engineer on the One AI team here at One Model.

We're going to dive into cross validation in machine learning models.

We will cover what cross validation is, why cross validation is important, cross validation and model evaluation, and how cross validation works in One AI.

After completing this module, you will understand the significance of cross validation in assessing the performance of predictive models on new and unseen data.

You will recognize the critical role of cross validation in preventing overfitting.

You will appreciate how cross validation provides a more accurate estimate of a model's performance by simulating training and testing on different subsets of the data. And finally, you will learn how cross validation works in One AI.

Chapter 2

Overview of Cross Validation

I will now provide an overview of cross validation from a top level.

Cross validation is a technique used to assess the performance of a predictive model on new and unseen data. It involves dividing the data set into multiple subsets called folds. One fold is always held out as a validation fold in order to remain new and unseen from the model.

The model is then trained on the rest of the folds, and the performance is assessed based on how well the model performs on the data in the validation fold, which is called the validation step.

This process is repeated multiple times with several different combinations of training and validation folds to ensure robustness and reliability of the model's performance metrics.

Then the performance results from each validation step are averaged to produce a more powerful estimate of the model's performance.

This helps us ensure that the model we select for deployment is strong and generalizes well to new data since we want our models to work well when new data is added versus being overfit to the data that it was trained upon.

Chapter 3

Importance of Cross Validation

Now I'm going to talk about the importance of cross validation. The main purpose of cross validation is to prevent overfitting. This occurs when a model is trained too well on the training data, capturing noise and irrelevant patterns in addition to the underlying relationships that we actually want it to learn. So while it performs well on the training data, it performs poorly on new and unseen data, making it a bit of a flop.

When we use cross validation, we evaluate the model on multiple subsets of data, ensuring that the model's performance is not overly influenced by specific patterns in the training data, making it quite the opposite of a flop. Cross validation allows for the maximum utilization of available data for both training and testing purposes. It ensures that the data points are used for both training and validation at some point during the process, leading to a more robust model evaluation.

Cross validation is also critical in evaluation of model performance. It provides a more accurate estimate of how well the model will generalize to new and unseen data by simulating the process of training and testing on different subsets of the data.

This is how we get to F1 holdout, precision holdout, and recall holdout scores in the result summary report each time a model is run. These scores help us know if we should trust the model and share it out, or if we need to finetune to produce a stronger, more performant model. Thanks to cross validation, you don't have to run your models on old data where we already know what happened to check how good they are.

Chapter 4

Cross Validation in One AI

I will now talk about how cross validation works in One AI. Cross validation is automatically used by One AI with preset configurations.

We have chosen a default number of five folds, which has been empirically demonstrated by several sources to yield low bias and variance.

However, if you think more or less folds would improve the validation scores, you can configure the number of folds in the advanced configuration settings through the One AI augmentation screen.

Chapter 7

Conclusion & Thanks

Thank you for joining me to learn about cross validation and machine learning models! You should now be more comfortable with how cross validation works and how it helps us both improve and assess the performance of machine learning models. We'll catch you next time. Happy modeling!