Lab 3.4 - Training a Machine Learning Model

Lab overview

In this lab, you will continue exploring the biomechanical vertebral column dataset. You will first split the dataset into three separate datasets for training, validation, and testing. You will then use this data to train a machine learning (MML) model by using the XGBoost algorithm.

Objectives

After completing this lab, you will be able to:

- Split data into training, validation and test datasets
- Train a XGBoost model in Amazon SageMaker

Prerequisites

This lab requires:

- Access to a notebook computer with Wi-Fi and Microsoft Windows, macOS, or Linux (Ubuntu, SUSE, or Red Hat)
- For Microsoft Windows users: Administrator access to the computer
- An internet browser such as Chrome, Firefox, or IE9 (previous versions of Internet Explorer are not supported)

Duration

This lab requires **30** minutes to complete. The lab will remain active for **120** minutes

Accessing the AWS Management Console

1. At the top of these instructions, choose Start Lab to launch your lab.
   A **Start Lab** panel opens, which displays the lab status.
2. Wait until you see the message *Lab status: ready*, then close the **Start Lab** panel by choosing the **X**.
3. At the top of these instructions, choose AWS
   This will open the AWS Management Console in a new browser tab. The system will automatically log you in.
   Tip: If a new browser tab does not open, there will typically be a banner or icon at the top of your browser that indicates that your browser is preventing the website from opening pop-up windows. Select the banner or icon and then choose **Allow pop ups**.
4. Arrange the **AWS Management Console** browser tab so that it displays next to these instructions. Ideally, you should be able to see both browser tabs at the same time, which can make it easier to follow the lab steps.

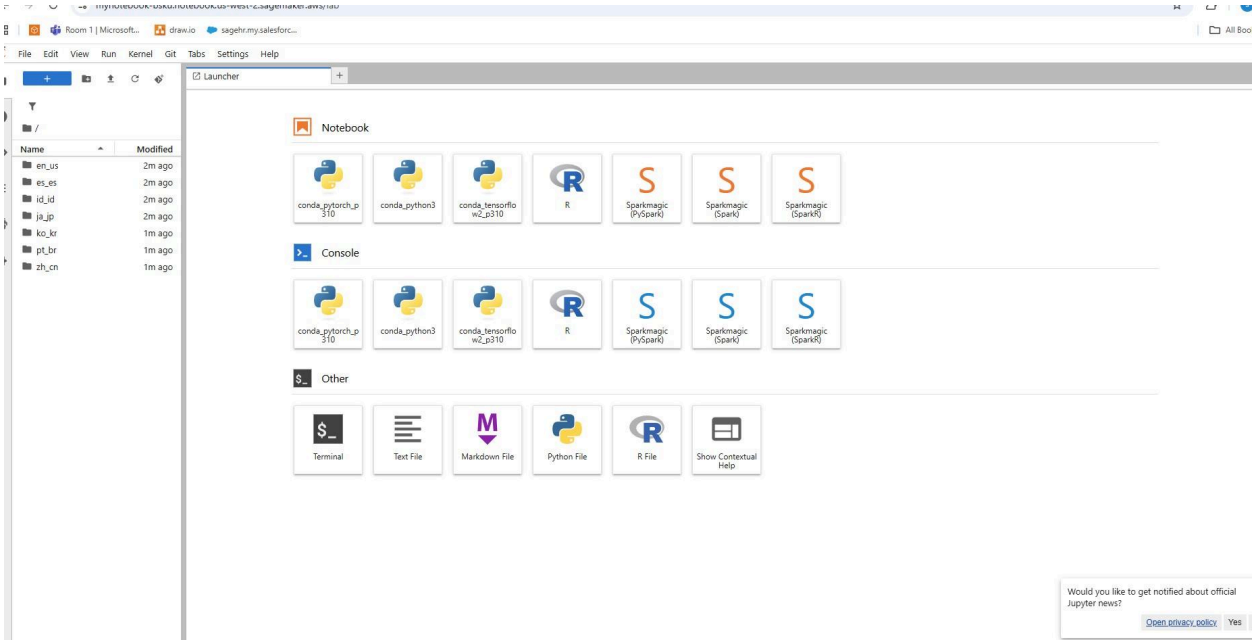
Task 1: Accessing a notebook instance in Amazon SageMaker

In this task, you will open your JupyterLab environment and switch to the notebook to complete the lab.

To open JupyterLab:

5. At the top of the AWS Management Console, in the search bar, search for and choose `Amazon SageMaker AI`.

6. From the navigation menu on the left, expand the **Applications and IDEs** section and choose **Notebooks** and then choose **Notebook instances** tab from lower pane.
7. Look for the notebook instance named `MyNotebook`. Open the JupyterLab notebook instance by going to the end of the row and choosing **Open JupyterLab**.



Task 2: Opening a notebook in your notebook instance

In this task, you will open the notebook for this lab:

8. In your JupyterLab environment, go to the file browser in the left pane and locate the `3_4-machinelearning.ipynb` file.
9. Open the `en_us/3_4-machinelearning.ipynb` file by choosing it.
   Tip: If a window appears asking you to select a kernel, select **conda_python3** and then choose **Select**.
10. Follow the instructions in the notebook.

Conclusion

You now have successfully:

- Split data into training, validation, and test datasets
- Trained an XGBoost model in Amazon SageMaker