

Lab 3.3 - Encoding Categorical Variables

Lab overview

In this lab, you will encode categorical variables. You will use the [automobile dataset](#) from the UC Irvine Machine Learning Repository. This dataset works well because it contains many categorical features.

Objectives

After completing this lab, you will be able to:

- Encode ordinal categorical data
- Encode non-ordinal categorical data

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 takes approximately **30 minutes**. The lab will remain active for **120 minutes**

AWS services not used in this lab

AWS services that are not used in this lab are disabled in the lab environment. In addition, the capabilities of the services used in this lab are limited to what the lab requires. Expect errors when you access other services or perform actions beyond the ones that are provided in this lab guide.

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. On the AWS Management Console, on the **Services** menu, choose **Amazon SageMaker**.
6. From the navigation menu on the left, expand the **Notebook** section and choose **Notebook instances**.
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_3-machinelearning.jpynb` file.
9. Open the `en_us/3_3-machinelearning.jpynb` file by choosing it.

Tip: If a window appears asking you to select a kernel, select **conda_python3**, then choose **Select**.

10. Follow the instructions in the notebook.

Conclusion

You now have successfully:

- Encoded ordinal categorical data
- Encoded non-ordinal categorical data

Lab complete

Congratulations! You have completed the lab.

11. To confirm that you want to end the lab, at the top of this page, choose End Lab, and then choose Yes.

A panel should appear with this message: *DELETE has been initiated... You may close this message box now.*

12. To close the panel, choose the **X** in the top-right corner.