

Lesson 2:
Building a Model using SageMaker

SEARCH

RESOURCES

CONCEPTS

✓

1. Introduction to Amazon SageMaker

✓

2. AWS Setup Instructions for Regula...

✓

3. AWS SageMaker

✓

4. SageMaker Instance Utilization L...

5. Setting up a Notebook Instance

6. Cloning the Deployment Notebooks

7. Is Everything Set Up?

8. Boston Housing Example - Getting...

9. Boston Housing Example - Trainin...

10. Boston Housing Example - Testin...

11. Mini-Project: Building Your First ...

12. Mini-Project: Solution

13. Boston Housing In-Depth - Data ...

14. Boston Housing In-Depth - Creati...

15. Boston Housing In-Depth - Buildi...

16. Boston Housing In-Depth - Creati...

17. Summary

A. AWS Service Utilization Quota (Limits)

You need to understand the way AWS imposes *utilization quotas* (limits) on almost all of its services. *Quotas*, also referred to as *limits*, are the maximum number of resources of a particular service that you can create in your AWS account.

- AWS provides default quotas, **for each AWS service**.
- Importantly, **each quota is region-specific**.
- There are three ways to **view your quotas**, as mentioned [here](#):
 1. Service Endpoints and Quotas,
 2. Service Quotas console, and
 3. AWS CLI commands - `list-service-quotas` and `list-aws-default-service-quotas`
- In general, there are three ways to **increase the quotas**:
 1. Using [Amazon Service Quotas](#) service - This service consolidates your account-specific values for quotas across all AWS services for improved manageability. Service Quotas is available at no additional charge. You can directly try logging into [Service Quotas console](#) here.
 2. Using [AWS Support Center](#) - You can create a case for support from AWS.
 3. AWS CLI commands - `request-service-quota-increase`

A.1. Amazon SageMaker Utilization Quota (Limits)

You can view the *Amazon SageMaker Service Limits* at "[Amazon SageMaker Endpoints and Quotas](#)" page. You can request to increase the AWS Sagemaker quota using the [AWS Support Center](#) only. Note that **currently the Amazon Service Quotas does not support SageMaker service**. However, SageMaker would be introduced soon into Service Quotas. AWS is moving to make users manage quotas for all AWS services from one central location.

Introducing Service Quotas: View and manage your quotas for AWS services from one central location

Posted On: Jun 24, 2019

With Service Quotas, you can view and manage your quotas easily and at scale as your AWS workloads grow. Quotas, also referred to as limits, are the maximum number of resources that you can create in an AWS account. AWS implements quotas to provide highly available and reliable service to all customers, and protect you from unintentional spend. Each quota starts with an AWS default value. Based on your needs, you can request to increase quota values for your specific account.

Service Quotas consolidates the AWS default values and your account specific values for quotas across AWS services in one single location, providing you with improved visibility. At launch, you can view default quotas for over 90 AWS services, with more coming soon.

Service Quotas makes the process of requesting quota increases easier. You simply search for a quota and put in your desired value to submit a quota increase request. You can proactively manage your quotas by configuring Amazon CloudWatch alarms that monitor usage and alert you to approaching quotas.

SageMaker would be introduced soon into Services Quota - Courtesy - [Amazon Service Quotas](#)

A.2. Increase Sagemaker Instance Quota (Limit) using AWS Support Center

Read the note and recommendation below before proceeding further.

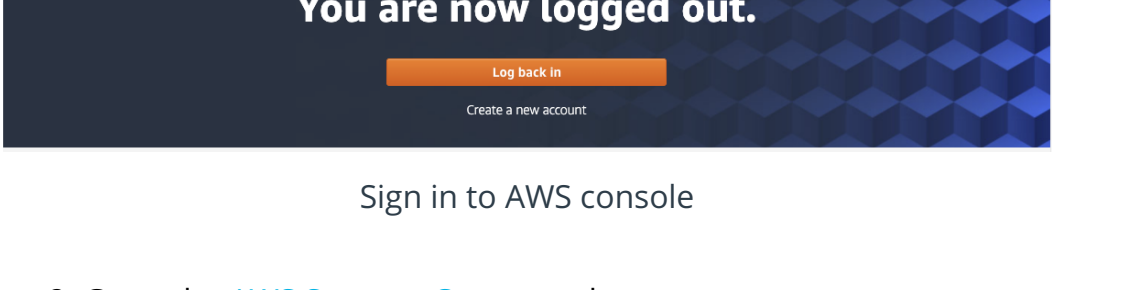
Note

Suppose a student has a quota of 20 instances of `ml.m4.xlarge` by default, they would not notice it unless they run the notebook that uses that instance. Now, if they go to the AWS Support Center, to request a service limit increase by 1, their instance limit will be degraded from 20 to 1.

Recommendation

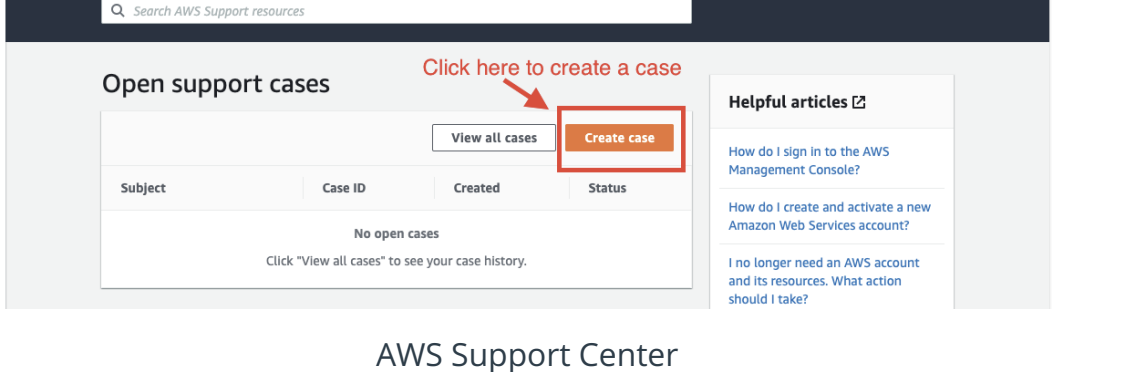
1. For `ml.m4.xlarge` - The default quota would be any number in the range [0 - 20]. Students can expect an error - *'ResourceLimitExceeded'*, when executing the notebook in the concept *Boston Housing Problem - Training The Model*, later in this lesson. In such a case only, the student must request a limit increase for `ml.m4.xlarge`.
2. For `ml.p2.xlarge` - The default quota would be either 0 or 1, therefore it is alright to go ahead and request an increase anytime.

1. Sign in to AWS console - <https://aws.amazon.com/console/>



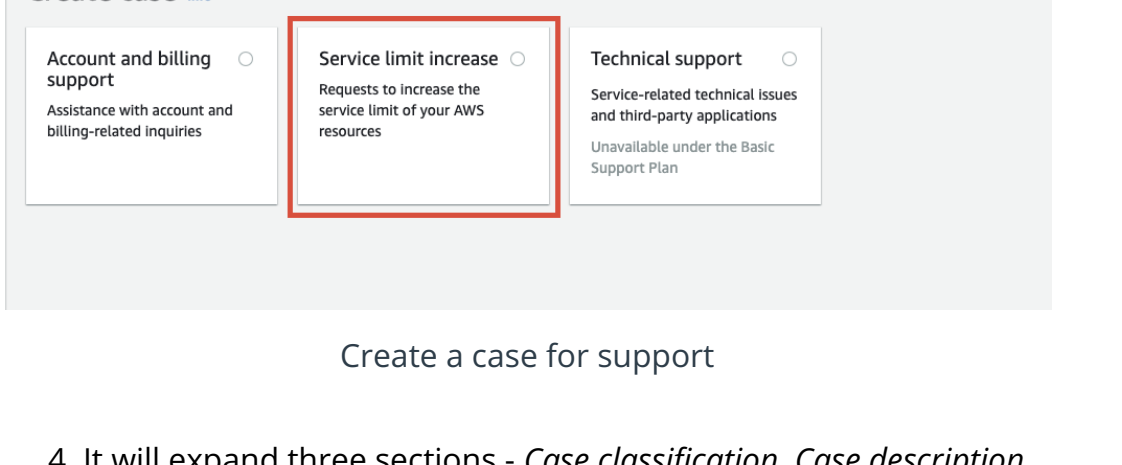
Sign in to AWS console

2. Go to the [AWS Support Center](#) and create a *case*.



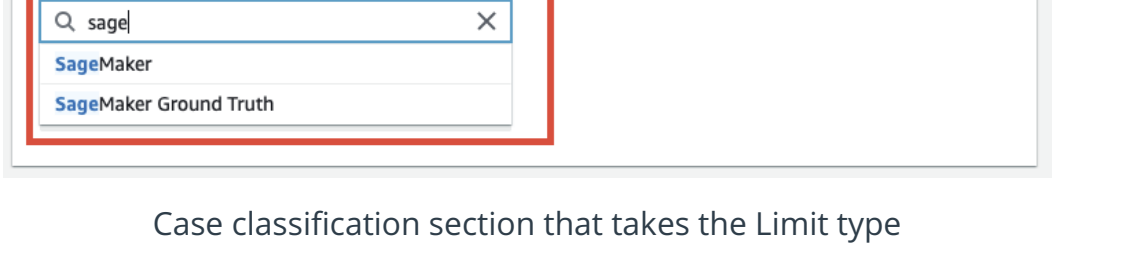
AWS Support Center

3. Click on *Service limit increase*



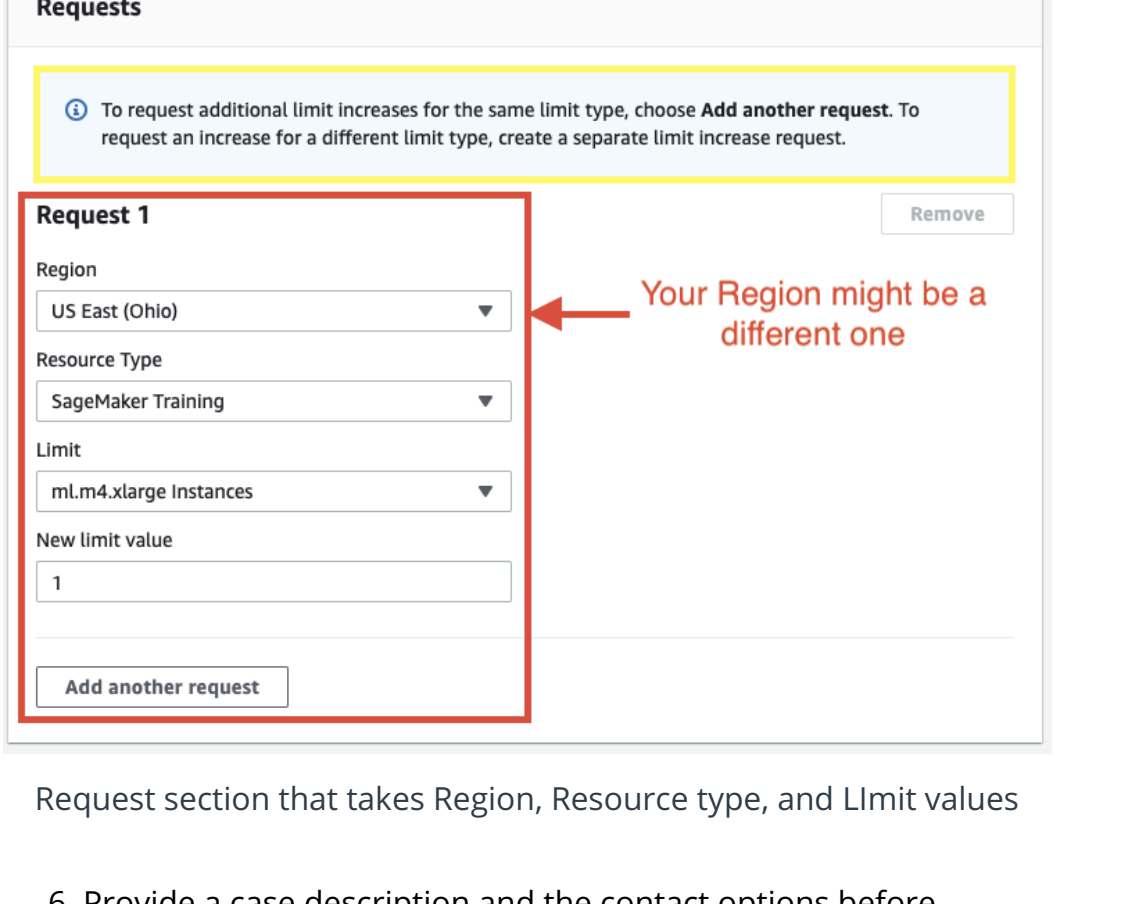
Create a case for support

4. It will expand three sections - *Case classification*, *Case description*, and *Contact options* on the same page. In *Case classification* section, select **"Sagemaker"** as the *Limit type*.



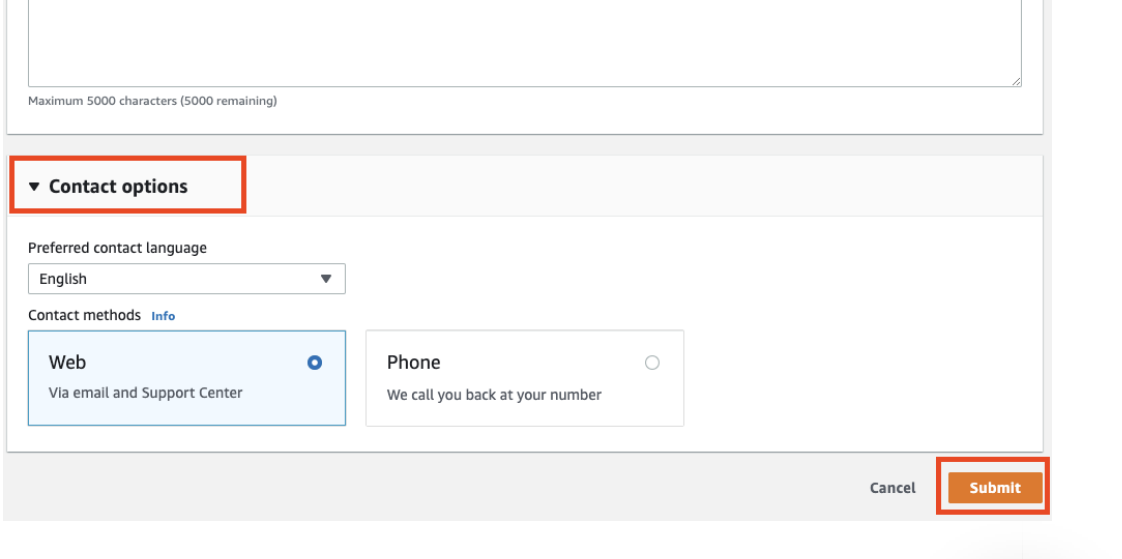
Case classification section that takes the Limit type

5. It will expand one more section - *Requests* on the same page. In *Request* section, and select the Region in which you are using the SageMaker service.
 - Select Sagemaker Training as the Resource Type
 - Select the instance type (ml.m4.xlarge or ml.p2.xlarge) under the Limit field
 - Under new limit value, select 1



Request section that takes Region, Resource type, and Limit values

6. Provide a case description and the contact options before submitting the case to support.



IMPORTANT NOTICE: *This is the current AWS UI as of April 6th, 2020. The AV UI is subject to change on a regular basis. We advise students to refer to AV documentation for the above process.*