

CS 193.1

Natural Language Processing & Amazon Web Services



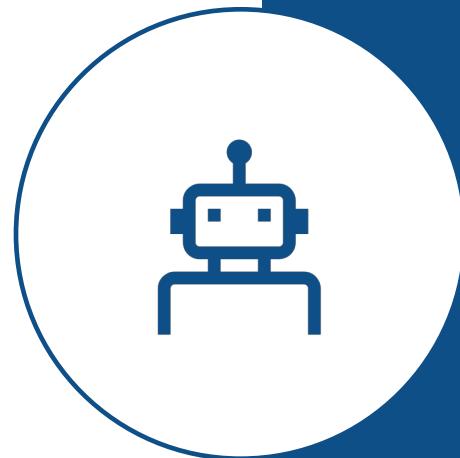
What is NLP?

A context-aware method of text mining.

NLP is king.

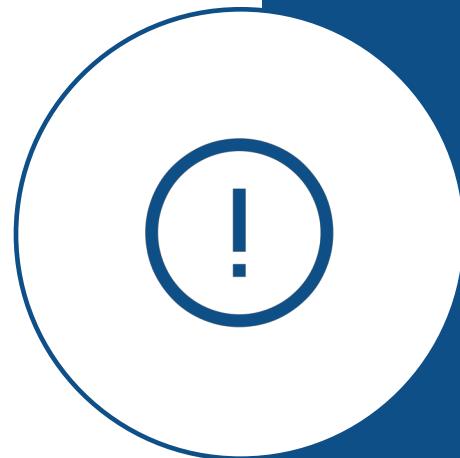
NLP includes methods and algorithms that take unstructured input into natural language data. This data can be a variety of things:

1. Parts of speech
2. Word relations and dependencies
3. Semantics and grammar
4. A lot more



Language is highly ambiguous and variable.

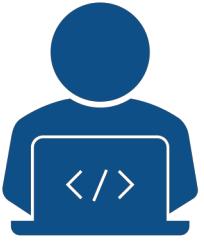
- **Ambiguous:**
 - “I ate pizza with friends”
 - “I ate pizza with olives”
- **Variable:**
 - “I ate pizza with friends”
 - “My friends and I ate pizza.”



What do we know about language?

- Humans are **great at using language** (producing statements and understanding language), but **not at understanding the formal rules** that govern language.
 - Computers are the opposite: they can perfectly understand formal rules but cannot innately disambiguate the nuances of language.
 - Calculating the shortest path in a graph vs. detecting sarcasm.



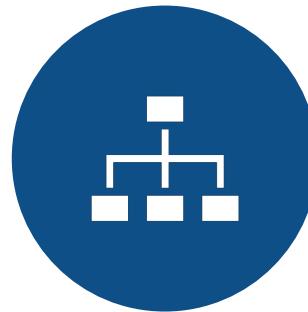


Goal of NLP: transform language into a computer-operable form that can still encapsulate meaning and context.



What is difficult in NLP?

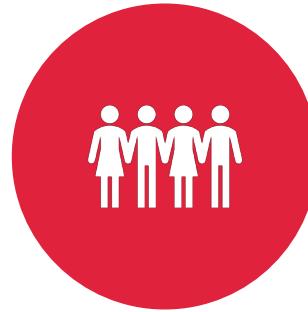
Linguistic challenges



Syntactic ambiguity
Multiplicity of statement structures



Semantic ambiguity
Multiplicity of statement definitions



Contextual variability
eras, groups, cultures



Human idiosyncrasy
slang, sarcasm, humor

What is difficult in NLP?

Technical challenges



Language is compositional: letters → words
→ phrases → sentence



There is more meaning as the composition becomes more complex.



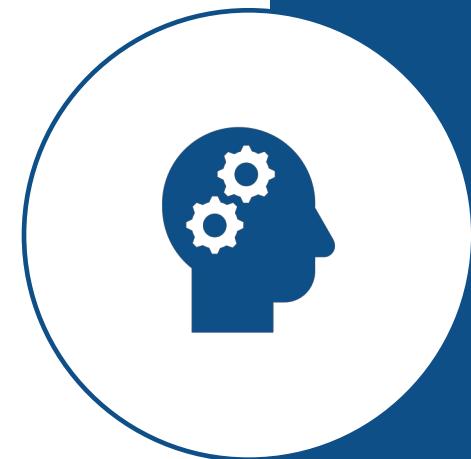
Data is sparse: the number of possible combinations of terms to form meaning is practically infinite.



Given this, our models may not be able to generalize over all our inputs.

What are common methods found in NLP?

- Linguistic features
 - Parts of speech (POS) tagging
 - Syntactic parsing (dependency parsing)
 - Segmentation, tokenization, lemmatization
- Supervised machine learning
 - Named entity recognition
 - Text classification
- Unsupervised machine learning
 - Information extraction
 - Text summarization



Amazon Web Services

Using the AWS Console, S3 buckets, and Sagemaker notebooks



Teach Tomorrow's Cloud Workforce Today

With the increasing demand for cloud employees, AWS Educate provides an academic gateway for the next generation of IT and cloud professionals. AWS Educate is Amazon's global initiative to provide students and educators with the resources needed to accelerate cloud-related learning.

[Join AWS Educate](#)

[Login to AWS Educate](#)

Join Hundreds of Thousands of Students, over 10,000 Educators, and More
Than 2,400 Institutions Who Use AWS Educate



hello :)

jann.montalan@obf.ateneo.edu



.....



SIGN IN A blue rectangular button with the text "SIGN IN" and a circular arrow icon to its right, both in white. The entire button is bordered by a red line.

[Forgot password?](#)

[Not an AWS Educate member? Apply now.](#)

**Jann Railey Montalan**Consecutive Days: **1**Pathways Completed: **1**Badges Earned: **1**

Preferred Language:

English



Career Path:

Cloud Computing 101

Digital certificate earned**COMPLETED**

Career Path:

Data Integration Specialist

Your Progress: 0 of 5 complete

CONTINUE LEARNING

Career Path:

Data Scientist



Suggested Jobs

Cloud Support Associate

Amazon, Inc.

[more about this opportunity](#)

Cloud Support Engineer

Amazon, Inc.

[more about this opportunity](#)

Cloud Support Associate

Amazon, Inc.

[more about this opportunity](#)

Cloud Support Engineer

Amazon, Inc.

[more about this opportunity](#)



AWS Educate Starter Account

Your cloud journey has only just begun. Use your AWS Educate Starter Account to access the AWS Console and resources, and start building in the cloud!

[AWS Educate Starter Account](#)

Your account has an estimated **75** credits remaining and access will end on **Apr 7, 2020**.

Note: Clicking this button will take you to a third party site managed by Vocareum, Inc. ("Third Party Servicer"). In addition to the AWS Educate terms of service, your use of the AWS Educate Starter Account is governed by the Third Party Servicer's terms, including its Privacy Policy. AWS assumes no responsibility or liability and makes no representations or warranties regarding services provided by a Third Party Servicer.

Welcome to AWS Educate Starter Account

Use your Starter Account to access to a wide variety of AWS Services and start building! Click on the AWS Console button to sign and get started.

- [What AWS services can I use in my Starter Account?](#)

You can use the following services in your Starter Account: apigateway, athena, cloud9, cloudformation, cloudfront, cloudtrail, cloudwatch, codecommit, codedeploy, codepipeline, cognito-identity, cognito-idp, cognito-sync, comprehend, deeplens, dynamodb, ec2, ecs, elasticache, elasticfilesystem, elasticloadbalancing, elasticmapreduce, events, execute-api, glue, iam, inspector, iot, kinesis, kinesisanalytics, firehose, kms, lambda, lex, logs, machinelearning, mobilehub, opsworks, polly, rds, rekognition, route53 (other than domain name purchasing), s3, sns, sqs, swf, sagemaker, translate, transcribe

- [What regions can I use with a Starter Account?](#)

- [Are Service Linked Roles supported?](#)

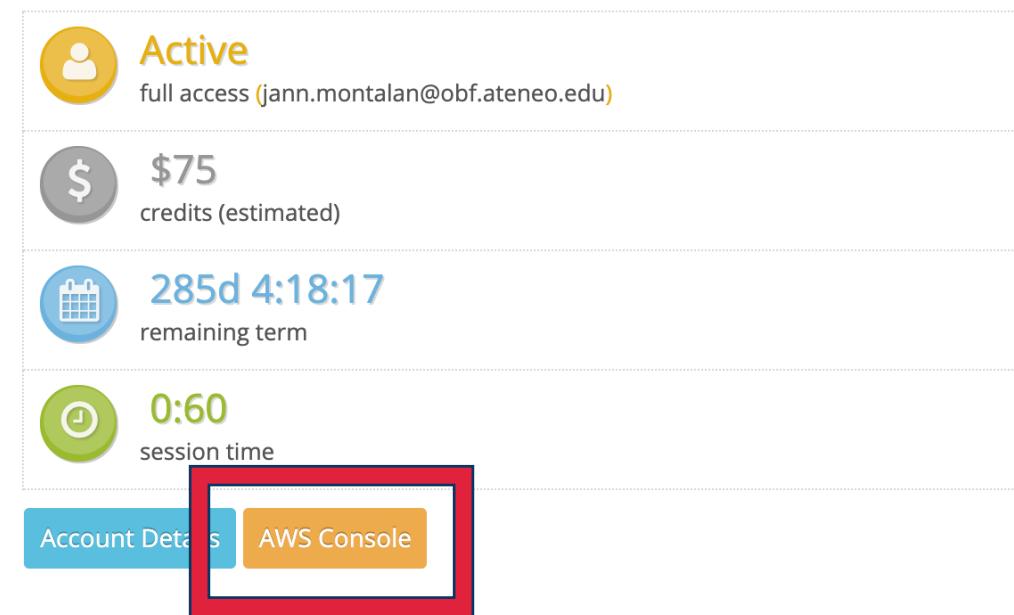
- [I can't start any resources. What happened?](#)

- [Can I create users within my Starter Account for others to access?](#)

- [Can I create my own IAM policy within Starter Account or Classroom?](#)

- [How can I use IAM roles within AWS services?](#)

Your Starter Account Status





AWS Management Console

AWS services

Find Services

You can enter names, keywords or acronyms.



Example: Relational Database Service, database, RDS

▼ Recently visited services

Amazon SageMaker

IAM

AWS Organizations

S3

Billing

► All services

Build a solution

Get started with simple wizards and automated workflows.

Launch a virtual machine

With EC2

2-3 minutes



Build a web app

With Elastic Beanstalk

6 minutes



Build using virtual servers

With Lightsail

1-2 minutes



Access resources on the go



Access the Management Console using the AWS Console Mobile App. [Learn more](#)

Explore AWS

Data Lake Storage

Build your data lake on the most secure, durable, and scalable storage. [Learn more](#)

Visit AWS around the world at a Summit

AWS Global Summits bring the cloud computing community together to connect, collaborate, and learn about AWS. [Learn more](#)

Open Distro for Elasticsearch

A 100% open-source, community driven distribution of Elasticsearch with enterprise-grade security and alerting features [Learn more](#)



Services ▾

Resource Groups ▾



vocstartsoft/user271338=jann.... ▾

N. Virginia ▾

Support ▾

History

Console Home

Amazon SageMaker

S3

IAM

Billing

AWS Organizations

iam

IAM

Manage User Access and Encryption Keys

ECS

Lightsail ↗

ECR

ECS

EKS

Lambda

Batch

Elastic Beanstalk

Serverless Application Repository

AWS RoboMaker

Blockchain

Amazon Managed Blockchain



Satellite

Ground Station

Athena

EMR

CloudSearch

Elasticsearch Service

Kinesis

QuickSight ↗

Data Pipeline

AWS Glue

MSK

Alexa for Business

Amazon Chime ↗

WorkMail

End User Computing

WorkSpaces

AppStream 2.0

WorkDocs

WorkLink

Internet Of Things

IoT Core

Amazon FreeRTOS

IoT 1-Click

IoT Analytics

IoT Device Defender

IoT Device Management

IoT Events

IoT Greengrass

IoT SiteWise

IoT Things Graph

Storage

S3

EFS

FSx

S3 Glacier

Storage Gateway

AWS Backup

Database

RDS

DynamoDB

ElastiCache

Nimongo

Management & Governance

AWS Organizations

CloudWatch

AWS Auto Scaling

CloudFormation

CloudTrail

Config

OpsWorks

Service Catalog

Systems Manager

Trusted Advisor

Managed Services

Control Tower

AWS License Manager

Security, Identity, & Compliance

IAM

Resource Access Manager

Cognito

Secrets Manager

GuardDuty

Inspector

Amazon Macie ↗

AWS Single Sign-On

Certificate Manager

Key Management Service

CloudHSM

Directory Service

▲ close

features Learn more ↗

Identity and Access Management (IAM)

AWS Account (968395985724)

[Dashboard](#)[Groups](#)[Users](#)[Roles](#)[Policies](#)[Identity providers](#)[Account settings](#)[Credential report](#)[Encryption keys](#) [Search IAM](#)

AWS Organizations

[Organization activity](#)[Service control policies \(SCPs\)](#)

Roles

What are IAM roles?

IAM roles are a secure way to grant permissions to entities that you trust. Examples of entities include the following:

- IAM user in another account
- Application code running on an EC2 instance that needs to perform actions on AWS resources
- An AWS service that needs to act on resources in your account to provide its features
- Users from a corporate directory who use identity federation with SAML

IAM roles issue keys that are valid for short durations, making them a more secure way to grant access.

Additional resources:

- [IAM Roles FAQ](#)
- [IAM Roles Documentation](#)
- [Tutorial: Setting Up Cross Account Access](#)
- [Common Scenarios for Roles](#)

[Create role](#)[Delete role](#)

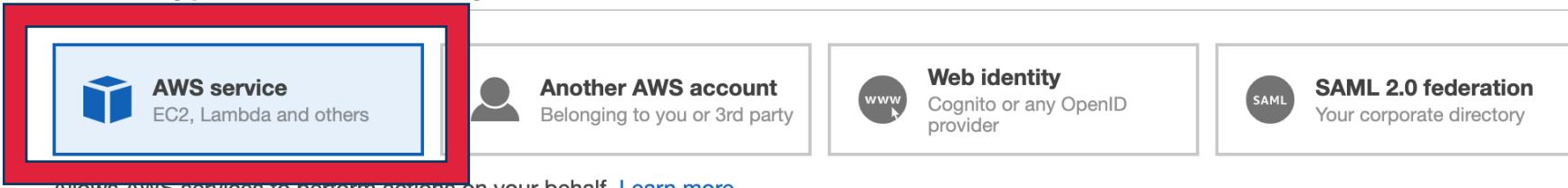
Showing 14 results

Role name ▾	Description	Trusted entities
<input type="checkbox"/> AWSServiceRoleForAWSCloud9		AWS service: cloud9 (Service-Linked role)
<input type="checkbox"/> AWSServiceRoleForCloudWatc...		AWS service: events (Service-Linked role)
<input type="checkbox"/> AWSServiceRoleForElastiCache		AWS service: elasticache (Service-Linked role)

Create role

1 2 3 4

Select type of trusted entity



Choose the service that will use this role

EC2

Allows EC2 instances to call AWS services on your behalf.

Lambda

Allows Lambda functions to call AWS services on your behalf.

API Gateway	Comprehend	EMR	Kinesis	S3
AWS Backup	Config	ElastiCache	Lambda	SMS
AWS Support	Connect	Elastic Beanstalk	Lex	SNS
Amplify	DMS	Elastic Container Service	License Manager	SWF
AppSync	Data Lifecycle Manager	Elastic Transcoder	Machine Learning	SageMaker
Application Auto Scaling	Data Pipeline	Elastic Load Balancing	Macie	
Application Discovery Service	DataSync	Forecast	MediaConvert	Service Catalog
Batch	DeepLens	Glue	OpsWorks	Step Functions
	Directory Service	Greengrass	Personalize	Storage Gateway

* Required

Cancel

Next: Permissions

API Gateway	Comprehend	EMR	Kinesis	S3
AWS Backup	Config	ElastiCache	Lambda	SMS
AWS Support	Connect	Elastic Beanstalk	Lex	SNS
Amplify	DMS	Elastic Container Service	License Manager	SWF
AppSync	Data Lifecycle Manager	Elastic Transcoder	Machine Learning	SageMaker
Application Auto Scaling	Data Pipeline	Elastic Load Balancing	Macie	Security Hub
Application Discovery Service	DataSync	Forecast	MediaConvert	Service Catalog
Batch	DeepLens	Glue	OpsWorks	Step Functions
CloudFormation	Directory Service	Greengrass	Personalize	Storage Gateway
CloudHSM	DynamoDB	GuardDuty	RAM	Transfer
CloudTrail	EC2	Inspector	RDS	Trusted Advisor
CloudWatch Application Insights	EC2 - Fleet	IoT	Redshift	VPC
CloudWatch Events	EC2 Auto Scaling	IoT Things Graph	Rekognition	WorkLink
CodeBuild	EKS	KMS	RoboMaker	WorkMail
CodeDeploy				

Select your use case

SageMaker - Execution

Allows SageMaker notebook instances, training jobs, and models to access S3, ECR, and CloudWatch on your behalf.

* Required

Cancel

Next: Permissions

Create role

1

2

3

4

Review

Provide the required information below and review this role before you create it.

Role name *

social_computing_sagemaker

Role description

Allows SageMaker notebook instances, training jobs, and models to access S3, ECR, and CloudWatch on your behalf.

Maximum 1000 characters. Use alphanumeric and '+=,.@-_' characters.

Trusted entities AWS service: sagemaker.amazonaws.com

Policies

 AmazonSageMakerFullAccess 

Permissions boundary Permissions boundary is not set

No tags were added.

* Required

Cancel

Previous

Create role

Identity and Access Management (IAM)

▼ AWS Account (968395985724)

[Dashboard](#)[Groups](#)[Users](#)

Roles

[Policies](#)[Identity providers](#)[Account settings](#)[Credential report](#)[Encryption keys](#) [Search IAM](#)

▼ AWS Organizations

[Organization activity](#)[Service control policies \(SCPs\)](#)[Create role](#)[Delete role](#)

Search

Showing 14 results

Role name	Description	Trusted entities
AWSServiceRoleForAWSCloud9		AWS service: cloud9 (Service-Linked role)
AWSServiceRoleForCloudWatc...		AWS service: events (Service-Linked role)
AWSServiceRoleForElastiCache		AWS service: elasticache (Service-Linked role)
AWSServiceRoleForOrganizations	Service-linked role used by AWS Organizations to enable integrati...	AWS service: organizations (Service-Linked r...
AWSServiceRoleForSupport	Enables resource access for AWS to provide billing, administrativ...	AWS service: support (Service-Linked role)
AWSServiceRoleForTrustedAdv...	Access for the AWS Trusted Advisor Service to help reduce cost, i...	AWS service: trustedadvisor (Service-Linked ...)
EMR_AutoScaling_DefaultRole		AWS service: elasticmapreduce and 1 more
EMR_DefaultRole		AWS service: elasticmapreduce
EMR_EC2_DefaultRole		AWS service: ec2
social-computing-sagemaker	Allows SageMaker notebook instances, training jobs, and models...	AWS service: rekognition and 2 more
vocareum		AWS service: sagemaker
vocstarhard		Account: 933987587310
vocstartsoft		Account: 933987587310

Identity and Access Management (IAM)

▼ AWS Account (968395985724)

Dashboard

Groups

Users

Roles

Policies

Identity providers

Account settings

Credential report

Encryption keys

 Search IAM

▼ AWS Organizations

Organization activity

Service control policies (SCPs)

Roles > social-computing-sagemaker

Summary

[Delete role](#)**Role ARN** arn:aws:iam::968395985724:role/social-computing-sagemaker [Copy](#)**Role description** Allows SageMaker notebook instances, training jobs, and models to access S3, ECR, and CloudWatch on your behalf. | [Edit](#)**Instance Profile ARNs** [Copy](#)**Path** /**Creation time** 2019-06-27 05:24 UTC+0800**Maximum CLI/API session duration** 1 hour [Edit](#)[Permissions](#)[Trust relationships](#)[Tags](#)[Access Advisor](#)[Revoke sessions](#)

▼ Permissions policies (2 policies applied)

[Attach policies](#)[+ Add inline policy](#)[Policy name ▾](#)[Policy type ▾](#)▶  AmazonS3FullAccess

AWS managed policy

[X](#)▶  AmazonSageMakerFullAccess

AWS managed policy

[X](#)

▶ Permissions boundary (not set)

Add permissions to social-computing-sagemaker

Attach Permissions

Create policy↻

Filter policies ▾🔍 s3Showing 3 results

Policy name	Type	Used as	Description
<input type="checkbox"/> ➔ AmazonDMSRedshiftS3Role	AWS managed	None	Provides access to manage S3 settings for Redshift endpoints for DMS.
<input type="checkbox"/> ➔ AmazonS3ReadOnlyAccess	AWS managed	None	Provides read only access to all buckets via the AWS Management Console.
<input type="checkbox"/> ➔ QuickSightAccessForS3Storage...	AWS managed	None	Policy used by QuickSight team to access customer data produced by S3 S...

CancelAttach policy

Amazon S3

[Buckets](#)[Batch operations](#)[Block public access
\(account settings\)](#)[Feature spotlight](#) 2

Welcome to Amazon S3. Create new buckets or select an existing bucket to view and configure properties.

[Documentation](#)

S3 buckets

[Discover the console](#) Search for buckets

All access types ▾

[+ Create bucket](#)[Edit public access settings](#)[Empty](#)[Delete](#)

1 Buckets

1 Regions

 Bucket name ▾Access [i](#) ▾

Region ▾

Date created ▾

 socialcomputing

Bucket and objects not public

Asia Pacific (Singapore)

Jun 27, 2019 4:56:37 AM
GMT+0800

Create bucket



① Name and region

② Configure options

③ Set permissions

④ Review

Name and region

Bucket name

social-computing

Region

Asia Pacific (Singapore)

Copy settings from an existing bucket

Select bucket (optional) 1 Buckets

Create

Cancel

Next

Amazon SageMaker



Dashboard

[Search](#)

Notebook

[Notebook instances](#)[Lifecycle configurations](#)[Git repositories](#)

Training

[Algorithms](#)[Training jobs](#)[Hyperparameter tuning jobs](#)

Inference

[Model packages](#)[Models](#)[Endpoint configurations](#)[Endpoints](#)[Batch transform jobs](#)[AWS Marketplace](#) 

Amazon SageMaker introduces Batch Transform

You can now run predictions on batch data with Batch Transform using a model in SageMaker. You pay only for the resources used during the job execution.

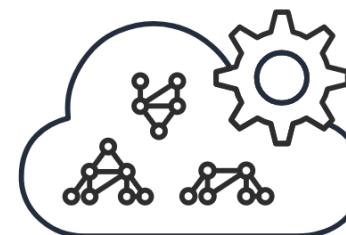
[Learn more](#)

Overview

[Hide](#)

Notebook

Availability of AWS and SageMaker SDKs and [sample notebooks](#) to create training jobs and deploy models.

[Notebook instances](#)

Training

Train and tune models at any scale. Leverage high performance AWS algorithms or bring your own.

[Training jobs](#)

Inference

Create models from training jobs or import external models for hosting to run inferences on new data.

[Models](#)

Recent activity

Recent activity within the

Last 7 days

Amazon SageMaker X

Dashboard

Search

▼ Notebook

[Notebook instances](#)

Lifecycle configurations

Git repositories

▼ Training

Algorithms

Training jobs

Hyperparameter tuning jobs

▼ Inference

Model packages

Models

Endpoint configurations

Endpoints

Batch transform jobs

AWS Marketplace 

Amazon SageMaker > Notebook instances

Notebook instances

Actions ▾

Create notebook instance

 Search notebook instances

Name	Instance	Creation time	Status	Actions
 social-computing	ml.t2.medium	Jun 26, 2019 21:19 UTC	 InService	Open Jupyter Open JupyterLab

Create notebook instance

Amazon SageMaker provides pre-built fully managed notebook instances that run Jupyter notebooks. The notebook instances include example code for common model training and hosting exercises. [Learn more](#) ↗

Notebook instance settings

Notebook instance name

Maximum length 63 characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

Notebook instance type



► Additional configuration

Permissions and encryption

IAM role

Notebook instances require permissions to call other services including SageMaker and S3. Choose a role or let us create a role with the [AmazonSageMakerFullAccess](#) IAM policy attached.



Root access - optional

- Enable - Give users root access to the notebook
 - Disable - Don't give users root access to the notebook
- Lifecycle configurations always have root access

Identity and Access Management (IAM)

AWS Account (968395985724)

[Dashboard](#)[Groups](#)[Users](#)

Roles

[Policies](#)[Identity providers](#)[Account settings](#)[Credential report](#)[Encryption keys](#) [Search IAM](#)

AWS Organizations

[Organization activity](#)[Service control policies \(SCPs\)](#)[Roles](#) > social-computing-sagemaker

Summary

[Delete role](#)

Role ARN

[arn:aws:iam::968395985724:role/social-computing-sagemaker](#)

Role permissions

[Access S3, ECR, and CloudWatch on your behalf.](#) | [Edit](#)

Instance Profile ARNs



Path

/

Creation time

2019-06-27 05:24 UTC+0800

Maximum CLI/API session duration

1 hour [Edit](#)

Permissions

Trust relationships

Tags

Access Advisor

Revoke sessions

▼ Permissions policies (2 policies applied)

[Attach policies](#)[+ Add inline policy](#)

	Policy name ▾	Policy type ▾	
▶	AmazonS3FullAccess	AWS managed policy	
▶	AmazonSageMakerFullAccess	AWS managed policy	

▶ Permissions boundary (not set)

Permissions and encryption

IAM Note

Notebook instances require permissions to call other services including SageMaker and Amazon S3. Choose a role or let us create a role with the [AmazonSageMakerFullAccess](#) IAM policy attached.

Enter a custom IAM role ARN

Custom IAM role ARN

arn:aws:iam::968395985724:role/social-computing-sagemaker

Choose a role or let us create a role with the



Enable - Give users root access to the notebook

Disable - Don't give users root access to the notebook

Lifecycle configurations always have root access

Encryption key - optional

Encrypt your notebook data. Choose an existing KMS key or enter a key's ARN.

No Custom Encryption

**▶ Network - optional****▶ Git repositories - optional****▶ Tags - optional**

Canc

Create notebook instance

Amazon SageMaker X

[Dashboard](#)[Search](#)

▼ Notebook

[Notebook instances](#)[Lifecycle configurations](#)[Git repositories](#)

▼ Training

[Algorithms](#)[Training jobs](#)[Hyperparameter tuning jobs](#)

▼ Inference

[Model packages](#)[Models](#)[Endpoint configurations](#)[Endpoints](#)[Batch transform jobs](#)[AWS Marketplace](#) ↗[Amazon SageMaker](#) > [Notebook instances](#)

Notebook instances

[Actions ▾](#)[Create notebook instance](#) Search notebook instances< 1 >

Name	Instance	Creation time	Status	Actions
social-computing	ml.t2.medium	Jun 26, 2019 21:19 UTC	InService	Open Jupyter Open JupyterLab

[Open Jupyter](#) [Open JupyterLab](#)

Files

Running

Clusters

SageMaker Examples

Conda

Select items to perform actions on them

Upload

New ▾

 0

Name ↴

Last Modified

File size

 ntM_20newsgroups_topic_modeling_2019-06-26

2 hours ago

[Open JupyterLab](#)[Quit](#)[xgboost_abalone.ipynb](#)[Preview](#)[Use](#)[xgboost_mnist.ipynb](#)[Preview](#)[Use](#)

Introduction to Applying Machine Learning

[Breast Cancer Prediction.ipynb](#)[Preview](#)[Use](#)[EnsembleLearnerCensusIncome.ipynb](#)[Preview](#)[Use](#)[fair_linear_learner.ipynb](#)[Preview](#)[Use](#)[gluon_recommender_system.ipynb](#)[Preview](#)[Use](#)[ntm_20newsgroups_topic_model.ipynb](#)[Preview](#)[Use](#)[sagemaker-countycensusclustering.ipynb](#)[Preview](#)[Use](#)[video-game-sales-xgboost.ipynb](#)[Preview](#)[Use](#)[xgboost_customer_churn.ipynb](#)[Preview](#)[Use](#)[xgboost_customer_churn_neo.ipynb](#)[Preview](#)[Use](#)[xgboost_direct_marketing_sagemaker.ipynb](#)[Preview](#)[Use](#)[Advanced Functionality](#)[Hyperparameter Tuning](#)[Reinforcement Learning](#)[Aws Marketplace](#)[Additional Examples Github Fastai Course V3](#)[Additional Examples Github Pytorch Tutorials](#)

An Introduction to SageMaker Neural Topic Model

Unsupervised representation learning and topic extraction using Neural Topic Model

1. [Introduction](#)
2. [Data Preparation](#)
3. [Model Training](#)
4. [Model Hosting and Inference](#)
5. [Model Exploration](#)

Introduction

Amazon SageMaker Neural Topic Model (NTM) is an unsupervised learning algorithm that attempts to describe a set of observations as a mixture of distinct categories. NTM is most commonly used to discover a user-specified number of topics shared by documents within a text corpus. Here each observation is a document, the features are the presence (or occurrence count) of each word, and the categories are the topics. Since the method is unsupervised, the topics are not specified upfront and are not guaranteed to align with how a human may naturally categorize documents. The topics are learned as a probability distribution over the words that occur in each document. Each document, in turn, is described as a mixture of topics.

In this notebook, we will use the Amazon SageMaker NTM algorithm to train a model on the [20NewsGroups](#) data set. This data set has been widely used as a topic modeling benchmark.

The main goals of this notebook are as follows:

1. learn how to obtain and store data for use in Amazon SageMaker,
2. create an AWS SageMaker training job on a data set to produce an NTM model,
3. use the model to perform inference with an Amazon SageMaker endpoint.
4. explore trained model and visualized learned topics

If you would like to know more please check out the [SageMaker Neural Topic Model Documentation](#).

Next time...

:

NLP and statistical analysis

Tool Exercise 2: NLP and AWS

Social relationships and processes