Next Steps

Amber Israelsen
AUTHOR | DEVELOPER | TRAINER

www.amberisraelsen.com

Course Overview

Course Introduction

Identifying Opportunities for Machine Learning

Defining Machine Learning Problems

Fetching and Preparing Data

Training and Evaluating the Model

Deploying and Monitoring the Model

The AWS Machine Learning Stack

Next Steps

Module Overview

Key points to remember for the exam
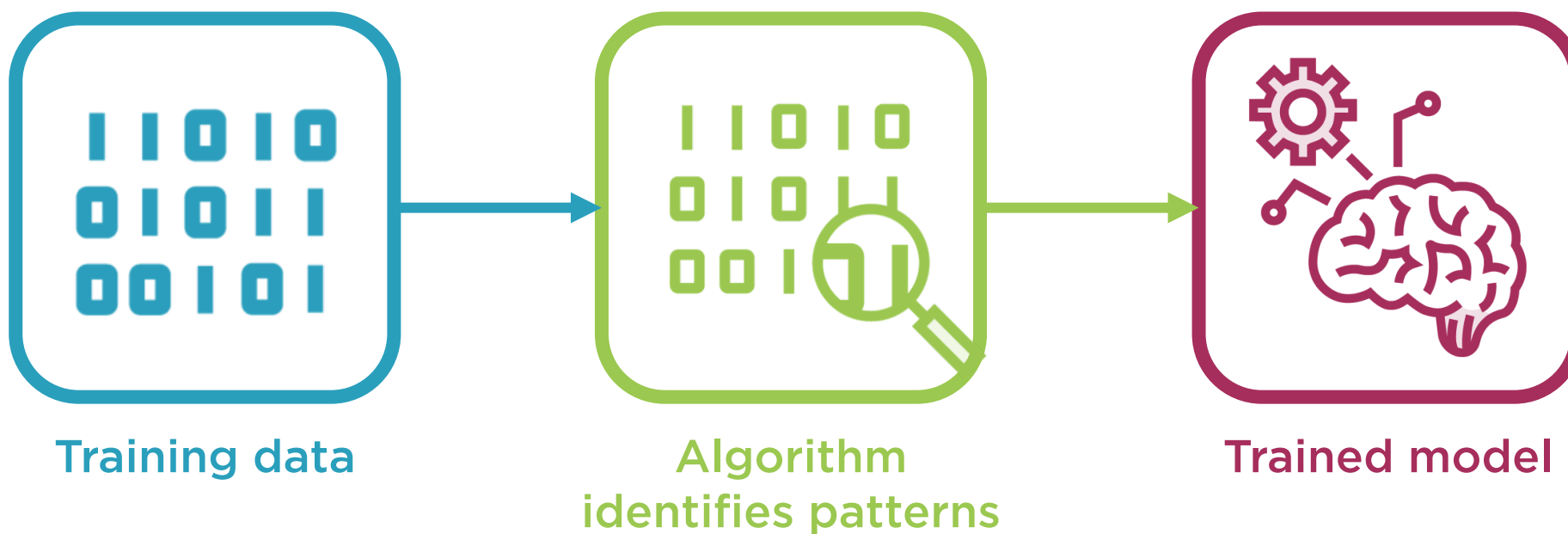
Additional resources

Wrapping up

Clean up your AWS resources!

S3 bucket
SageMaker endpoint
SageMaker application
Jupyter Notebook instance

Building a Definition of Machine Learning

Training data

**Algorithm
identifies patterns**

Trained model

Types of Machine Learning Problems

SUPERVISED

Learn from labeled data

Classification

BINARY

[yes/no]

[true/false]

[fraud, not fraud]

MULTICLASS

[cat, dog, horse]

[house, condo, townhome, apartment]

Regression

Uses continuous values

[predicting a stock price]

[predicting sales of a product]

UNSUPERVISED

Learn from finding hidden patterns in unlabeled data

Clustering

Groups data into clusters based on similar features

[after analyzing patient data, you find that certain groups respond better to treatment]

Anomaly Detection

Finds outliers in data

[suspicious network traffic]

[abnormal heart beat]

REINFORCEMENT

Learn from interacting with the environment to maximize reward

The algorithm receives feedback when it does something right or wrong

[gaming]

[recommendation systems]

[robot navigation]

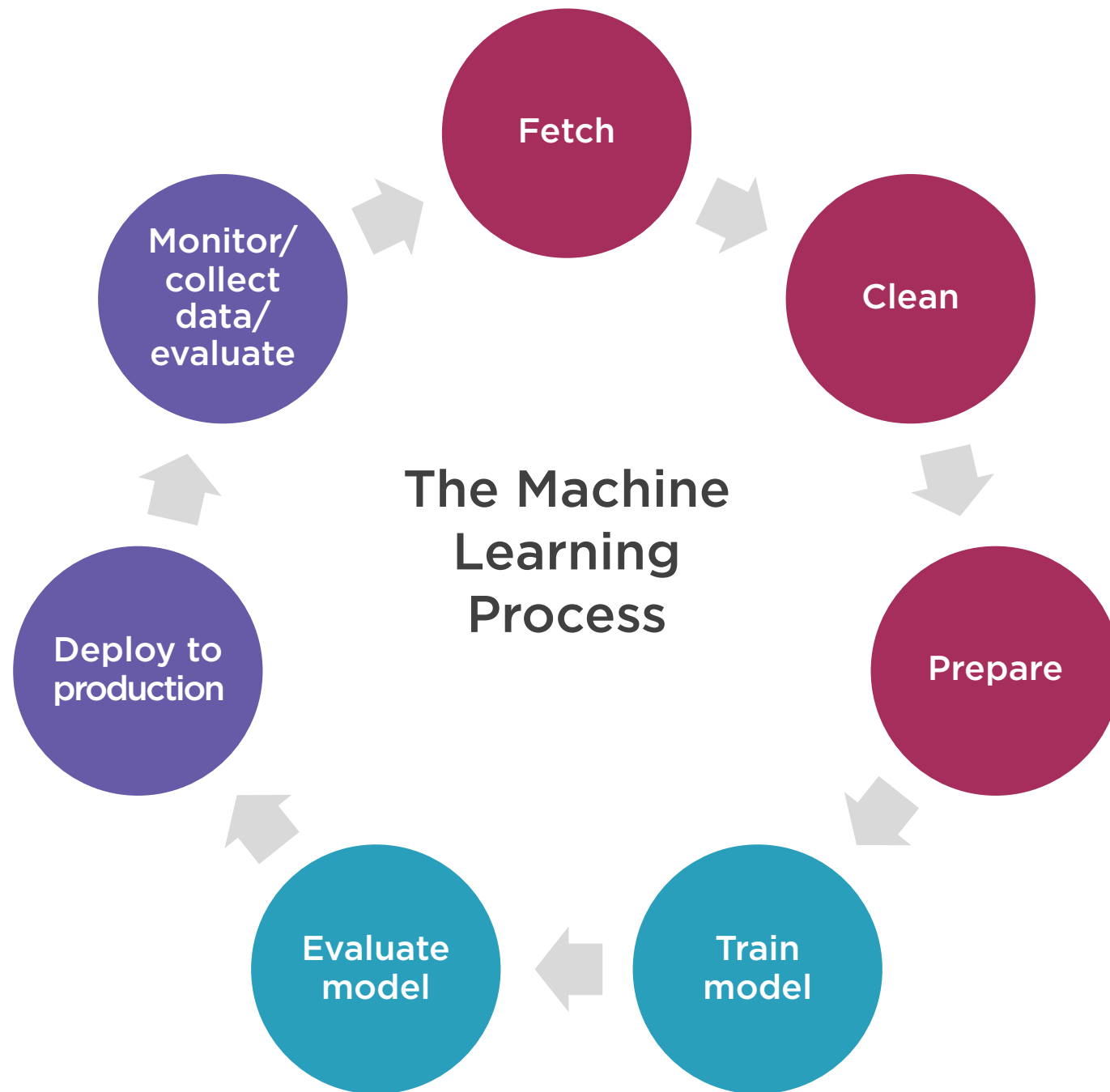[self-driving cars, like Amazon's Deep Racer]

Product Name	Price	Free Shipping	Release Date	Discount Eligible	Part of Bundle	Sold More Than 100 Units
3D globe puzzle	$29.95	Yes	12/2017	Yes	Yes	True
Passport holder	$20.45	Yes	2/2018	No	No	True
Women's travel pants	$47.88	No	4/2017	Yes	Yes	False
Scratch-it maps	$19.99	No	12/2018	No	Yes	False
Luggage tags	$12.99	Yes	3/2017	Yes	No	True
Monogrammed luggage	$129.79	No	2/2019	Yes	Yes	False
Men's travel vest	$39.99	No	11/2017	Yes	No	False

The Machine Learning Process

- Fetch
- Clean
- Prepare
- Train model
- Evaluate model
- Deploy to production
- Monitor/collect data/evaluate

Fetch

Batch
- Data Pipeline, Database Migration Service (DMS), AWS Glue

Streaming
- Kinesis family of services
- Data Streams, Data Firehose, Data Analytics, Video Streams

Amazon EMR, Athena, Redshift Spectrum

Clean and Prepare

Sanitize data to handle missing values, outliers, formatting and scale

Data visualizations
- Scatter plot, correlation matrix, histogram, box plots

Feature engineering

Ground Truth

Categorical Data
Describes Categories or Groups

NOMINAL
Order does Not matter

{Red, Yellow, Blue}

{Yes, No}

ORDINAL
Order does matter

{Small, Medium, Large}

{Hot, Hotter, Hottest}

One-hot Encoding

	Type
1	Truck
2	SUV
3	Sedan
4	Truck
5	Coupe

	Type_Truck	Type_SUV	Type_Sedan	Type_Coupe
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	1	0	0	0
5	0	0	0	1

Train and Evaluate

Algorithms

Data splitting

Steps to create a model

Evaluation

- Goal is to have the model generalize, not overfit or underfit
- Hyperparameter tuning
- Confusion matrix
- Accuracy, precision, recall

Deploy and Monitor

Use architecture and security best practices

Deploy a model
- Create the endpoint configuration first, then the endpoint

Monitoring
- CloudWatch, CloudTrail

Additional Resources

Additional Resources

Pluralsight

– Learning Path: *AWS Certified Machine Learning – Specialty*

AWS

– SageMaker Developer Guide

https://docs.aws.amazon.com/sagemaker/latest/dg/gs.html

– SageMaker Blog

https://aws.amazon.com/blogs/?awsf.blog-master-category=category%23artificial-intelligence

– re:Invent videos

– SageMaker Studio examples

– Certification home page

https://aws.amazon.com/certification/certified-machine-learning-specialty/

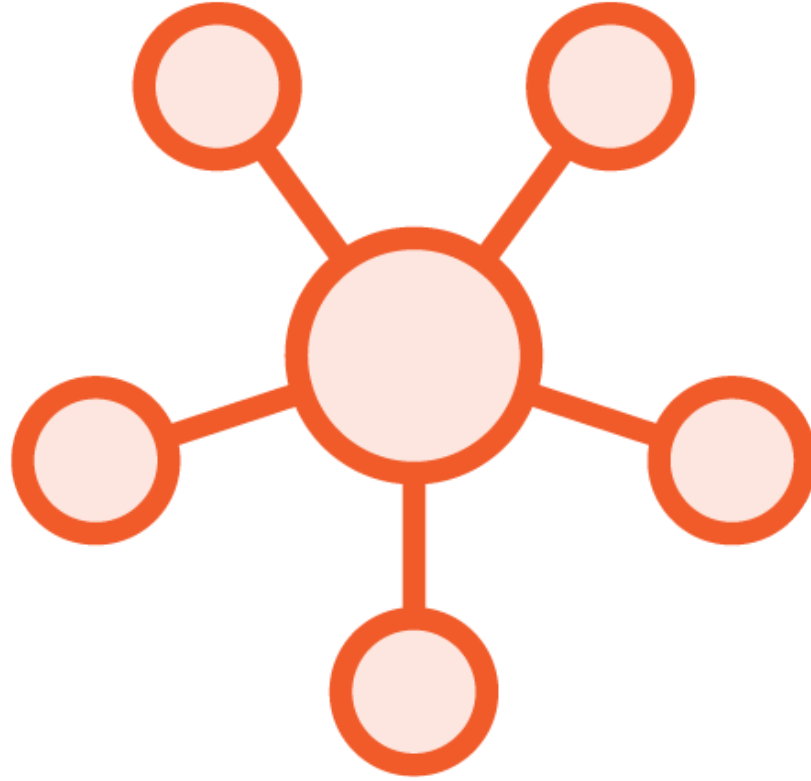- Consider taking the practice test

Wrapping Up

Let me hear from you!
Post on the discussion
board for this course.

Caring is sharing. Spread the word!

THANK YOU!