Exercise

Train and evaluate a regression model
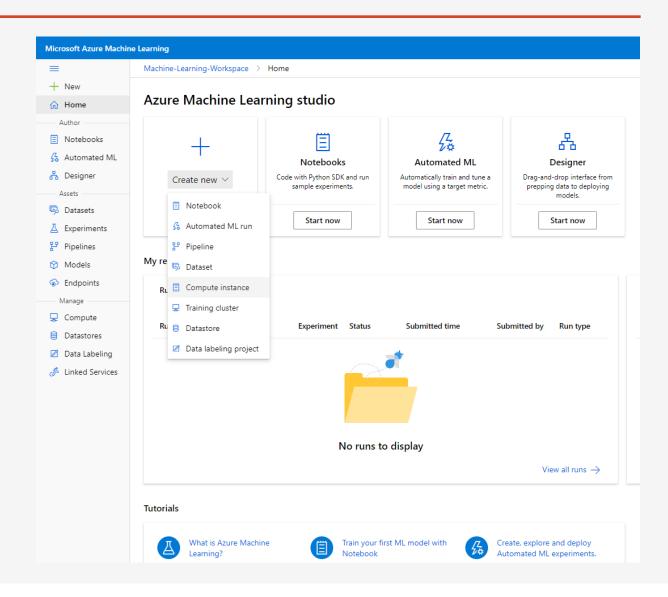
This exercise is based on the AI-900 exam material.

https://docs.microsoft.com/en-us/learn/certifications/exams/ai-900

Microsoft CERTIFIED

AZURE AI FUNDAMENTALS
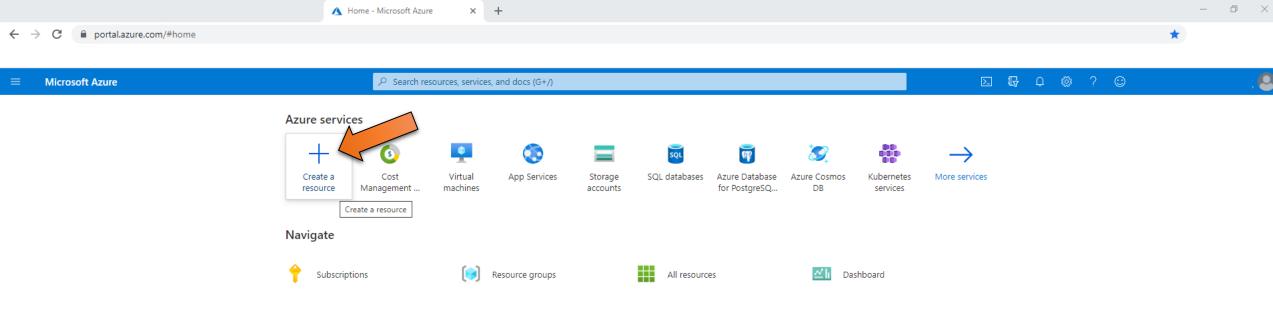
Train and evaluate a regression model

Calculating a regression line for a simple binomial (two-variable) function from first principles is possible but involves some mathematical effort. When you consider a real-world dataset in which x is not a single feature value such as temperature, but a vector of multiple variables such as temperature, day of week, month, rainfall, and so on; the calculations become more complex.

For this reason, data scientists generally use specialized machine learning frameworks to perform model training and evaluation. Such frameworks encapsulate common algorithms and provide useful functions for preparing data, fitting data to a model, and calculating model evaluation metrics.

Home - Microsoft Azure

portal.azure.com/#home

Microsoft Azure

Search resources, services, and docs (G+/)

Azure services

Create a resource

Cost Management ...

Virtual machines

App Services

Storage accounts

SQL databases

Azure Database for PostgreSQ...

Azure Cosmos DB

Kubernetes services

More services

Create a resource

Navigate

Subscriptions

Resource groups

All resources

Dashboard

Tools

Microsoft Learn
Learn Azure with free online training from Microsoft

Azure Monitor
Monitor your apps and infrastructure

Security Center
Secure your apps and infrastructure

Cost Management
Analyze and optimize your cloud spend for free

Useful links

Technical Documentation
Azure Migration Tools

Azure Services
Find an Azure expert

Recent Azure Updates
Quickstart Center

Azure mobile app

Download on the App Store

GET IT ON Google Play

https://portal.azure.com/#create/hub

portal.azure.com/#create/hub

Microsoft Azure

Search resources, services, and docs (G+/)

Home > New >

Machine Learning
Microsoft

Machine Learning
Microsoft

♥ Add to Favorites

★★★★½ 4.3 (105...)

Create

Overview   Plans   Usage Information + Support   Reviews

Azure Machine Learning empowers developers and data scientists with a wide range of productive experiences for building, training, and deploying machine learning models.
Create an Azure Machine Learning workspace to train, manage, and deploy machine-learning experiments and web services.

Media

More offers from Microsoft   See All

Workspace	Wire Data 2.0	Microsoft HPC Pack 2012 R2	Windows 10 IoT Core Services
Microsoft	Microsoft	Microsoft	Microsoft
No rating	★★★½☆ 3.5 (2 ratings)	★★★★☆ 3.9 (7 ratings)	★★★☆☆ 3.0 (2 ratings)
Virtual Machine	Azure Service	Virtual Machine	Azure Service
Windows Virtual Desktop resource	Provides the ability to explore wire data and helps identify network related issues.	Enterprise-class HPC solution. Easy to deploy, cost-effective and supports Windows/Linux workloads.	Commercialize your project with enterprise-grade security and support

https://portal.azure.com/#

portal.azure.com/#create/Microsoft.MachineLearningServices

Microsoft Azure

Search resources, services, and docs (G+/)

Home > New > Machine Learning >

Machine learning
Create a machine learning workspace

Basics   Networking   Advanced   Tags   Review + create

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ
Azure for Students

Resource group * ⓘ
unique-RG

Create new

Workspace details

Specify the name and region for the workspace.

Workspace name * ⓘ
Machine-Learning-Workspace

Region * ⓘ
East US 2

Storage account * ⓘ
(new) machinelearnin0583588374

Create new

Key vault * ⓘ
(new) machinelearnin7943657368

Create new

Application insights * ⓘ
(new) machinelearnin0858615732

Create new

Container registry * ⓘ
None

Create new

Review + create   < Previous   Next : Networking

Microsoft Azure

🔍 Search resources, services, and docs (G+/)

Home >

Microsoft.MachineLearningServices | Overview
Deployment

Delete   Cancel   Redeploy   Refresh

🚀 We'd love your feedback! →

⋯ Deployment is in progress

Deployment name: Microsoft.MachineLearningServices   Start time: 3/3/2021, 4:17:24 PM
Subscription: Azure for Students   Correlation ID: cc5ebde1-0d93-4c27-982e-00ee245ebed1
Resource group: unique-RG

∧ Deployment details (Download)

Resource	Type	Status	Operation details
🔄 machinelearnin7943657368	Microsoft.KeyVault/vaults	OK	Operation details
✅ machinelearnin0858615732	Microsoft.Insights/components	OK	Operation details
🔄 machinelearnin0583588374	Microsoft.Storage/storageAccounts	Accepted	Operation details

Overview
Inputs
Outputs
Template

Security Center
Secure your apps and infrastructure
Go to Azure security center >

Free Microsoft tutorials
Start learning today >

Work with an expert
Azure experts are service provider partners
who can help manage your assets on Azure
and be your first line of support.
Find an Azure expert >

portal.azure.com/#home

Microsoft Azure

Search resources, services, and docs (G+/)

Azure services

+
Create a
resource

All resources

All resources ★

+ Create   ⊙ View

Storage
accounts

SQL databases

Azure Database
for PostgreSQ...

Azure Cosmos
DB

→
More services

Navigate

🔑 Subscriptions

Resource groups

All resources

Dashboard

Tools

Microsoft Learn ↗
Learn Azure with free online
training from Microsoft

Azure Monitor
Monitor your apps and
infrastructure

Security Center
Secure your apps and
infrastructure

Cost Management
Analyze and optimize your
cloud spend for free

Useful links

Technical Documentation ↗
Azure Migration Tools

Azure Services ↗
Find an Azure expert

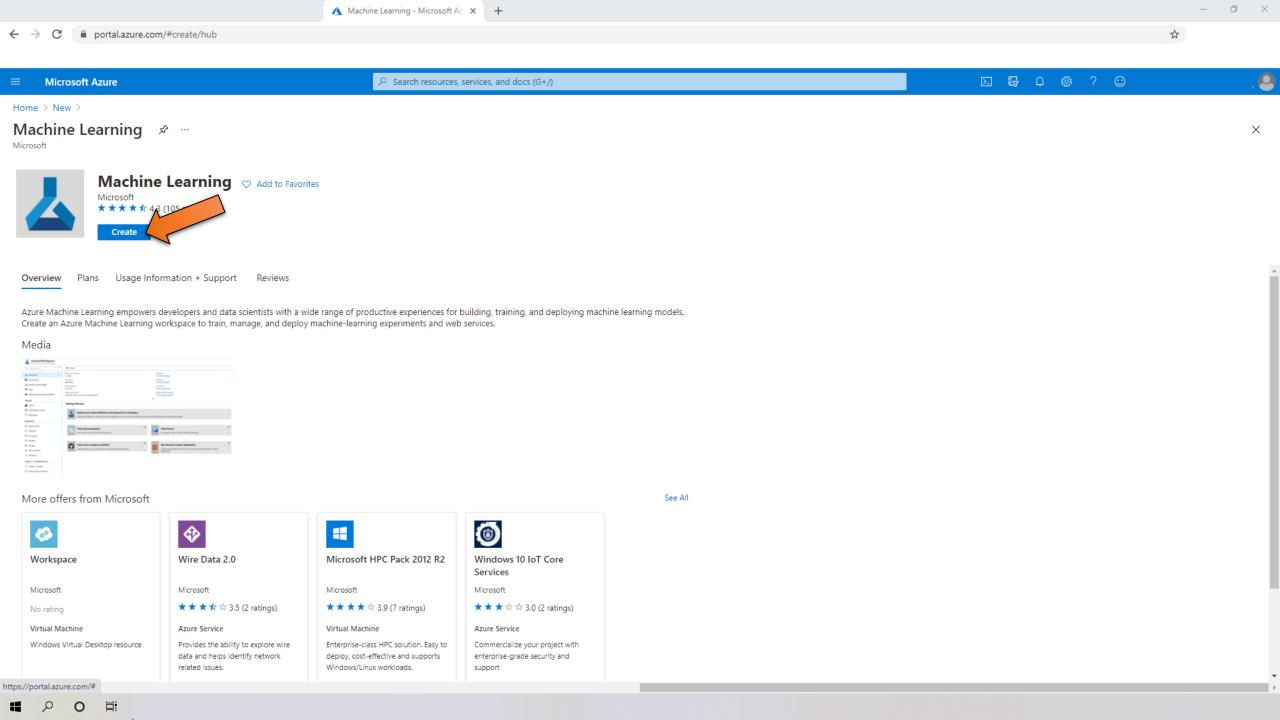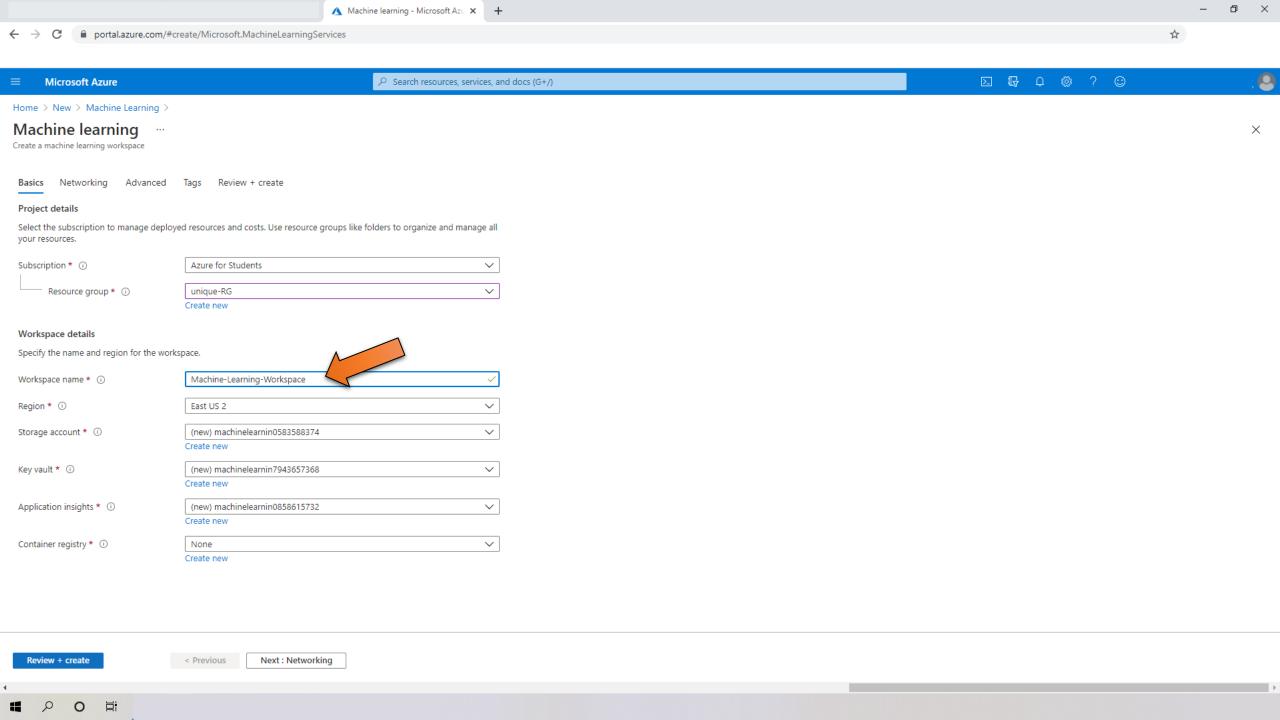Recent Azure Updates ↗
Quickstart Center

Azure mobile app

Download on the App Store

GET IT ON Google Play

https://portal.azure.com/#blade/HubsExtension/BrowseAll

All resources - Microsoft Azure

portal.azure.com/#blade/HubsExtension/BrowseAll

Microsoft Azure

Search resources, services, and docs (G+/)

Home >

All resources
Default Directory

+ Add   Manage view ∨   ↻ Refresh   ↓ Export to CSV   Open query   Assign tags   Delete   Feedback

Filter for any field...   Subscription == all   Resource group == all ✕   Type == all ✕   Location == all ✕   + Add filter

Showing 1 to 5 of 5 records.   ☐ Show hidden types ⓘ   No grouping ∨   List view ∨

☐ Name ↑↓	Type ↑↓	Resource group ↑↓	Location ↑↓	Subscription ↑↓	
☐ 🧪 Machine-Learning-Workspace	Machine learning	unique-RG	East US 2	Azure for Students	⋯
☐ 🗄 machinelearnin0583588374	Storage account	unique-RG	East US 2	Azure for Students	⋯
☐ 💡 machinelearnin0858615732	Application Insights	unique-RG	East US 2	Azure for Students	⋯
☐ 🔑 machinelearnin7943657368	Key vault	unique-RG	East US 2	Azure for Students	⋯
☐ 🌐 unique-RG	Cognitive Services	unique-RG	South Central US	Azure for Students	⋯

< Previous   Page 1 ∨ of 1   Next >

portal.azure.com/#@caiogasparinecagmail.onmicrosoft.com/resource/subscriptions/0a7a0cb1-eaca-4871-ad1d-b27ba6c8d60c/resourceGroups/unique-RG/providers/Microsoft.MachineLearningServices/workspaces/Machine-Learning-Workspace/overview

Microsoft Azure

Search resources, services, and docs (G+/)

Home > All resources >

All resources
Default Directory

+ Add   ⚙ Manage view ∨   ⋯

Filter for any field...

Machine-Learning-Workspace
Machine learning

Search (Ctrl+/)

⬇ Download config.json   🗑 Delete

⌄ Essentials   JSON View

Name ↑↓		
🧪 Machine-Learning-Workspace	⋯	
▦ machinelearnin0583588374	⋯	
🔑 machinelearnin0858615732	⋯	
💡 machinelearnin7943657368	⋯	
📦 unique-RG	⋯	

🧪 Overview
📋 Activity log
👥 Access control (IAM)
🏷 Tags
🔧 Diagnose and solve problems
⚡ Events

Settings
Private endpoint connections
Properties
🔒 Locks

Monitoring
Alerts
Metrics
Diagnostic settings
Logs

Automation
Tasks (preview)
Export template

Support + troubleshooting
Usage + quotas
New support request

Resource group : unique-RG
Location : East US 2
Subscription : Azure for Students
Subscription ID : 0a7a0cb1-eaca-4871-ad1d-b27ba6c8d60c

Studio web URL : https://ml.azure.com/?tid=1c4ea517-22ec-46e0-b2f4-8e3803dab82d&wsid=/su...
Storage : machinelearnin0583588374
Registry : ...
Key Vault : machinelearnin7943657368
Application Insights : machinelearnin0858615732

Manage your machine learning lifecycle

Use the Azure Machine Learning studio to build
evaluate, and deploy machine l...

Launch studio

Page [1 ∨] of 1

ml.azure.com/?tid=1c4ea517-22ec-46e0-b2f4-8e3803dab82d&wsid=/subscriptions/0a7a0cb1-eaca-4871-ad1d-b27ba6c8d60c/resourcegroups/unique-RG/workspaces/Machine-Learning-Workspace

Microsoft Azure Machine Learning

Machine-Learning-Workspace > Home

Azure Machine Learning studio

+ New

⌂ Home

Author

Notebooks

Automated ML

Designer

Assets

Datasets

Experiments

Pipelines

Models

Endpoints

Manage

Compute

Datastores

Data Labeling

Linked Services

+

Create new ⌄

Notebooks

Code with Python SDK and sample experiments.

Start now

Welcome to the studio!

Azure Machine Learning helps you build, train, deploy, and manage your models at cloud scale.

Register data

Reference data from storage to easily access during model training and explore using summary statistics.

Train models

Use machine learning algorithms with training data to create models.

Evaluate models

Find the best model using test data.

Deploy models

Deploy model as a web service in the Azure cloud, or to IoT Edge devices.

Start the tour

My recent resources

Runs

Run	Run ID	Experiment	S	Provisioning state	Created on

No runs to display

No Compute to display

View all runs →

View all compute →

Tutorials

What is Azure Machine Learning?

Train your first ML model with Notebook

Create, explore and deploy Automated ML experiments.

ml.azure.com/?tid=1c4ea517-22ec-46e0-b2f4-8e3803dab82d&wsid=/subscriptions/0a7a0cb1-eaca-4871-ad1d-b27ba6c8d60c/resourcegroups/unique-RG/workspaces/Machine-Learning-Workspace

Microsoft Azure Machine Learning

Machine-Learning-Workspace > Home

+ New

Home

Author

Notebooks

Automated ML

Designer

Assets

Datasets

Experiments

Pipelines

Models

Endpoints

Manage

Compute

Datastores

Data Labeling

Linked Services

Azure Machine Learning studio

+

Create new ⌄

	Notebooks	Automated ML	Designer
	Code with Python SDK and run sample experiments.	Automatically train and tune a model using a target metric.	Drag-and-drop interface from prepping data to deploying models.
	Start now	Start now	Start now

Notebook

Automated ML run

Pipeline

Dataset

Compute instance

Training cluster

Datastore

Data labeling project

My re

Ru

Ru   Experiment   Status   Submitted time   Submitted by   Run type

No runs to display

View all runs →

Compute

Name   Type   Provisioning state   Created on

No Compute to display

View all compute →

Tutorials

What is Azure Machine Learning?

Train your first ML model with Notebook

Create, explore and deploy Automated ML experiments.

ml.azure.com/compute/list/instances?wsid=/subscriptions/0a7a0cb1-eaca-4871-ad1d-b27ba6c8d60c/resourcegroups/unique-RG/workspaces/Machine-Learning-Workspace&tid=1c4ea517-22ec-46e0-b2f4-8e3803dab82d&createMode=true&computeType=...

Microsoft Azure Machine Learning

Machine-Learning-Workspace › Compute

Compute

Compute instances   Compute clusters   Inferer

ⓘ In the wake of COVID-19, we are prioritizing maintaining

Create compute instance

Select virtual machine

Select the virtual machine size you would like to use for your compute instance. Please note that only the creator can access the compute instance by default. Alternatively, you can assign the access to another user in the advanced settings section.

Location ⓘ

eastus2

Virtual machine type ⓘ

(●) CPU   ( ) GPU

Virtual machine size ⓘ

(●) Select from recommended options   ( ) Select from all options

Total available quota: 24 cores ⓘ

Name ↑	Category	Workload types	Available qu... ⓘ	Cost ⓘ
○ Standard_DS2_v2 2 cores, 7GB RAM, 14GB storage	General purpose	Development on Notebooks (or other IDE) and light weight testing	6 cores	$0.11/hr
● Standard_DS3_v2 4 cores, 14GB RAM, 28GB storage	General purpose	Classical ML model training, AutoML runs, pipeline runs (default compute)	6 cores	$0.23/hr
○ Standard_DS12_v2 4 cores, 28GB RAM, 56GB storage	Memory optimized	Training on large datasets (>1GB) parallel run steps, batch inferencing	6 cores	$0.30/hr
○ Standard_F4s_v2 4 cores, 8GB RAM, 32GB storage	Compute optimized	Real-time inferencing and other latency-sensitive tasks	6 cores	$0.17/hr

Back   Next   Cancel

○ Virtual Machine
○ Settings

ml.azure.com/compute/list/instances?wsid=/subscriptions/0a7a0cb1-eaca-4871-ad1d-b27ba6c8d60c/resourcegroups/unique-RG/workspaces/Machine-Learning-Workspace&tid=1c4ea517-22ec-46e0-b2f4-8e3803dab82d

Microsoft Azure Machine Learning

Machine-Learning-Workspace > Compute

Compute

Compute instances   Compute clusters   Inference clusters   Attached compute

ⓘ In the wake of COVID-19, we are prioritizing maintaining service availability for first responders, health and emergency management services, critical government infrastructure, and existing paid customer use. To balance the best possible experience for each of our customers, we have temporarily removed default vCPU/...

╋ New   ↻ Refresh   ▷ Start   ▢ Stop   ⟳ Restart   🗑 Delete   ▦ View quota | Instances I can use ⌄ | ⓘ   ⧩ Search to filter items...

Name	Status	Application URI	Virtual machine size	Created on ↓
Compute-name-9823a	◯ Creating		STANDARD_DS3_V2	Mar 3, 2021 4:23 PM

ml.azure.com/compute/list?wsid=/subscriptions/0a7a0cb1-eaca-4871-ad1d-b27ba6c8d60c/resourcegroups/unique-RG/workspaces/Machine-Learning-Workspace&tid=1c4ea517-22ec-46e0-b2f4-8e3803dab82d

Microsoft Azure Machine Learning

CG

Machine-Learning-Workspace > Compute

Compute

| Compute instances | Compute clusters | Inference clusters | Attached compute |

ⓘ In the wake of COVID-19, we are prioritizing maintaining service availability for first responders, health and emergency management services, critical government infrastructure, and existing paid customer use. To balance the best possible experience for each of our customers, we have temporarily removed default vCPU/... ✕

+ New   ↻ Refresh   ▶ Start   ◻ Stop   ⟳ Restart   🗑 Delete   ▦ View quota | Instances I can use ▾ | ⓘ   🔽 Search to filter items...

Name	Status	Application URI	Virtual machine size	Created on ↓
Compute-name-9823a	▶ Running	JupyterLab Jupyter RStudio Terminal	STANDARD_DS3_V2	Mar 3, 2021 4:23 PM

ml.azure.com/compute/list?wsid=/subscriptions/0a7a0cb1-eaca-4871-ad1d-b27ba6c8d60c/resourcegroups/unique-RG/workspaces/Machine-Learning-Workspace&tid=1c4ea517-22ec-46e0-b2f4-8e3803dab82d

Microsoft Azure Machine Learning

Machine-Learning-Workspace > Compute

Compute

Compute instances   Compute clusters   Inference clusters   Attached compute

ⓘ In the wake of COVID-19, we are prioritizing maintaining service availability for first responders, health and emergency management services, critical government infrastructure, and existing paid customer use. To balance the best possible experience for each of our customers, we have temporarily removed default vCPU/... ⌄   ✕

+ New   ⟳ Refresh   ▶ Start   ⏹ Stop   ⟳ Restart   🗑 Delete   ▤ View quota | Instances I can use ⌄ | ⓘ   ⚲ Search to filter items...

☑ **Name**   **Status**   **Application URI**   **Virtual machine size**   **Created on** ↓

☑ Compute-name-9823a   ▶ Running   JupyterLab Jupyter RStudio Termin

IMPORTANT NOTE: Always use trusted code   ✕

When using Azure Machine Learning, content within notebooks or
scripts that you load can potentially read data from your sessions and
access data within your organization in Azure. Only load notebooks or
scripts into Azure from trusted sources where you have reviewed the
source code. Learn more about secure best practices with Azure Machine
Learning here.

☑ Yes, I understand

Continue   Cancel

azureuser@compute-name-9823a:/mnt/batch/tasks/shared/LS_root/mounts/clusters/compute-name-9823a/code$
azureuser@compute-name-9823a:/mnt/batch/tasks/shared/LS_root/mounts/clusters/compute-name-9823a/code$
azureuser@compute-name-9823a:/mnt/batch/tasks/shared/LS_root/mounts/clusters/compute-name-9823a/code$
azureuser@compute-name-9823a:/mnt/batch/tasks/shared/LS_root/mounts/clusters/compute-name-9823a/code$
azureuser@compute-name-9823a:/mnt/batch/tasks/shared/LS_root/mounts/clusters/compute-name-9823a/code$ cd Users
azureuser@compute-name-9823a:/mnt/batch/tasks/shared/LS_root/mounts/clusters/compute-name-9823a/code/Users$ git clone https

cd Users
git clone https://github.com/caiogasparine/ml-basics

compute-name-9823a.eastus2.instances.azureml.ms/terminals/1

jupyter
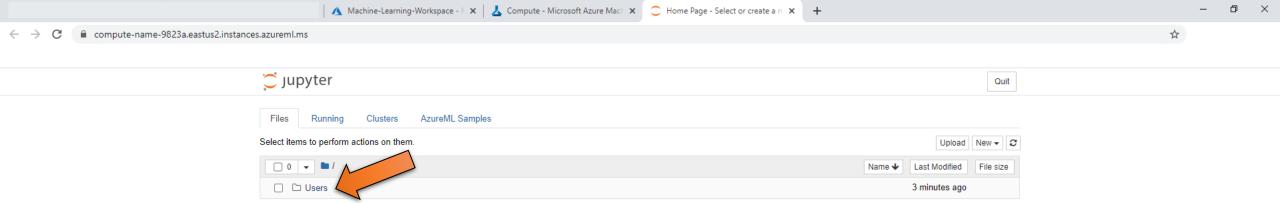
```
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.
azureuser@compute-name-9823a:/mnt/batch/tasks/shared/LS_root/mounts/clusters/compute-name-9823a/code$
azureuser@compute-name-9823a:/mnt/batch/tasks/shared/LS_root/mounts/clusters/compute-name-9823a/code$
azureuser@compute-name-9823a:/mnt/batch/tasks/shared/LS_root/mounts/clusters/compute-name-9823a/code$
azureuser@compute-name-9823a:/mnt/batch/tasks/shared/LS_root/mounts/clusters/compute-name-9823a/code$


Cloning into 'ml-basics'...
remote: Enumerating objects: 57, done.
remote: Counting objects: 100% (57/57), done.
remote: Compressing objects: 100% (50/50), done.
remote: Total 1943 (delta 27), reused 16 (delta 7), pack-reused 1886
Receiving objects: 100% (1943/1943), 8.28 MiB | 4.33 MiB/s, done.
Resolving deltas: 100% (726/726), done.
Updating files: 100% (1641/1641), done.
azureuser@compute-name-9823a:/mnt/batch/tasks/shared/LS_root/mounts/clusters/compute-name-9823a/code/Users$
```

compute-name-9823a.eastus2.instances.azureml.ms/tree/Users/ml-basics

📙 **jupyter**

Quit

Files   Running   Clusters   AzureML Samples

Select items to perform actions on them.

Upload   New ▾   ↻

☐ 0 ▾   📁 / Users / ml-basics   Name ↓   Last Modified   File size

📁 ..   seconds ago

☐ 📁 challenges   2 minutes ago

☐ 📁 data   a minute ago

☐ 📁 models   a minute ago

☐ 📖 01 - Data Exploration.ipynb   2 minutes ago   62.6 kB

☐ 📖 02 - Regression.ipynb   2 minutes ago   47.3 kB

☐ 📖 03 - Classification.ipynb   2 minutes ago   49 kB

☐ 📖 04 - Clustering.ipynb   2 minutes ago   15.2 kB

☐ 📖 05a - Deep Neural Networks (PyTorch).ipynb   2 minutes ago   23.8 kB

☐ 📖 05a - Deep Neural Networks (TensorFlow).ipynb   2 minutes ago   20.7 kB

☐ 📖 05b - Convolutional Neural Networks (PyTorch).ipynb   2 minutes ago   22.4 kB

☐ 📖 05b - Convolutional Neural Networks (Tensorflow).ipynb   2 minutes ago   16.4 kB

☐ 📖 05c - Transfer Learning (PyTorch).ipynb   2 minutes ago   20.4 kB

☐ 📖 05c - Transfer Learning (Tensorflow).ipynb   2 minutes ago   15.5 kB

☐ 📖 CODE_OF_CONDUCT.md   2 minutes ago   444 B

☐ 📖 local-setup.md   a minute ago   2.09 kB

☐ 📖 README.md   2 minutes ago   357 B

☐ 📖 SECURITY.md   2 minutes ago   2.78 kB

☐ 📄 LICENSE   2 minutes ago   1.14 kB

compute-name-9823a.eastus2.instances.azureml.ms/notebooks/Users/ml-basics/02%20-%20Regression.ipynb

Jupyter 02 - Regression (autosaved)

File　Edit　View　Insert　Cell　Kernel　Help

Kernel starting, please wait... | Trusted | Python 3.6 - AzureML ●

Markdown ▾

Regression

Supervised machine learning techniques involve training a model to operate on a set of *features* and predict a *label* using a dataset that includes some already-known label values. The training process *fits* the features to the known labels to define a general function that can be applied to new features for which the labels are unknown, and predict them. You can think of this function like this, in which y represents the label we want to predict and x represents the features the model uses to predict it.

$$y = f(x)$$

In most cases, x is actually a *vector* that consists of multiple feature values, so to be a little more precise, the function could be expressed like this:

$$y = f([x_1, x_2, x_3, \ldots])$$

The goal of training the model is to find a function that performs some kind of calculation to the x values that produces the result y. We do this by applying a machine learning *algorithm* that tries to fit the x values to a calculation that produces y reasonably accurately for all of the cases in the training dataset.

There are lots of machine learning algorithms for supervised learning, and they can be broadly divided into two types:

- ***Regression* algorithms**: Algorithms that predict a y value that is a numeric value, such as the price of a house or the number of sales transactions.
- ***Classification* algorithms**: Algorithms that predict to which category, or *class*, an observation belongs. The y value in a classification model is a vector of probability values between 0 and 1, one for each class, indicating the probability of the observation belonging to each class.

In this notebook, we'll focus on *regression*, using an example based on a real study in which data for a bicycle sharing scheme was collected and used to predict the number of rentals based on seasonality and weather conditions. We'll use a simplified version of the dataset from that study.

> **Citation**: The data used in this exercise is derived from Capital Bikeshare and is used in accordance with the published license agreement.

Explore the Data

The first step in any machine learning project is to explore the data that you will use to train a model. The goal of this exploration is to try to understand the relationships between its attributes; in particular, any apparent correlation between the *features* and the *label* your model will try to predict. This may require some work to detect and fix issues in the data (such as dealing with missing values, errors, or outlier values), deriving new feature columns by transforming or combining existing features (a process known as *feature engineering*), *normalizing* numeric features (values you can measure or count) so they're on a similar scale, and *encoding* categorical features (values that represent discrete categories) as numeric indicators.

Let's start by loading the bicycle sharing data as a **Pandas** DataFrame and viewing the first few rows.

In [ ]:
```
import pandas as pd

# load the training dataset
```

Machine-Learning-Workspace - | × | Compute - Microsoft Azure Mac | × | Users/ml-basics/ | × | 02 - Regression - Jupyter Notebo | ×

compute-name-9823a.eastus2.instances.azureml.ms/notebooks/Users/ml-basics/02%20-%20Regression.ipynb

jupyter 02 - Regression (unsaved changes)

File | Edit | View | Insert | Cell | Kernel | | Help

Trusted | Python 3.6 - AzureML ○

```python
for col in numeric_features:
    fig = plt.figure(figsize=(9, 6))
    ax = fig.gca()
    feature = bike_data[col]
    label = bike_data['rentals']
    correlation = feature.corr(label)
    plt.scatter(x=feature, y=label)
    plt.xlabel(col)
    plt.ylabel('Bike Rentals')
    ax.set_title('rentals vs ' + col + '- correlation: ' + str(correlation))
plt.show()
```

In [7]:

rentals vs temp- correlation: 0.5432846616821877

rentals vs atemp- correlation: 0.5438636902622049