

AzureML로 시작하는 Machine Learning

Microsoft Student Partner

KAIST 정태영

Session

- Session 1 : Introduction to Introduction to ML
머신러닝의 기초의 기초에 대해 배웁니다.
- Session 2 : Azure Machine Learning 맛보기
Azure ML studio를 통해 붓꽃을 분류하는
머신 러닝 프로그램을 만들어봅니다.

Azure ML?

- Microsoft에서 제공하는 ML 솔루션 클라우드
- 드래그&드롭
- Supervised / Unsupervised Learning 지원

예제 실습

- 붓꽃의 특성값을 이용해 종 분류하기
- 데이터의 구성은 다음과 같음
 - 1) 꽃받침의 길이
 - 2) 꽃받침의 너비
 - 3) 꽃잎의 길이
 - 4) 꽃잎의 너비

예제 실습



Iris setosa



Iris versicolor



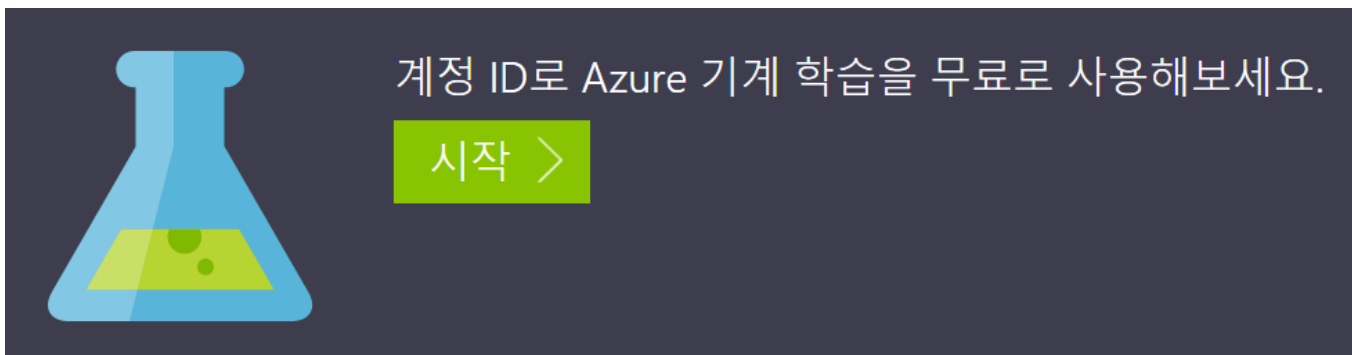
Iris virginica

Step 1 : 접속

- Azureml.com 접속
- 8-hour trial

Step 1 : 접속

- Azureml.com 접속
- 8-hour trial



Step 1 : 접속

- Azureml.com 접속
- 8-hour trial

Quick Evaluation	Most Popular	Enterprise Grade
Guest Workspace	Free Workspace	Standard Workspace
8-hour trial	\$0/month	\$9.99/month
No sign-in required.	Don't already have a Microsoft account? Simply sign up here .	Azure subscription required Other charges may apply. Read more .
Enter	Sign In	Create Workspace
<ul style="list-style-type: none">▪ No hassle instant access▪ Stock sample datasets▪ ML models built in minutes▪ Full range of ML algorithms	<ul style="list-style-type: none">▪ Free access that never expires▪ 10 GB storage on us▪ R and Python scripts support▪ Predictive web services	<ul style="list-style-type: none">▪ Full SLA Support▪ Bring your own Azure storage▪ Parallel graph execution▪ Elastic Web Service endpoints

Step 2 : 실험 생성

- 왼쪽 아래의 + new 버튼 클릭
- Blank Experiment

Step 2 : 실험 생성

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The top navigation bar includes the text "Microsoft Azure Machine Learning Studio", a timer showing "time remaining: 08:00", the user role "Guest", a "SIGN UP" button, and icons for help, team, and user profile.

The left sidebar contains a list of resources: PROJECTS, EXPERIMENTS (highlighted), NOTEBOOKS, DATASETS, and TRAINED MODELS. At the bottom of the sidebar is a "+ NEW" button, which is indicated by a large blue arrow.

The main content area is titled "experiments" and shows tabs for "MY EXPERIMENTS" and "SAMPLES". Below the tabs is a table with columns: NAME, AUTHOR, STATUS, LAST EDITED, and PROJECT. The table is currently empty, with the text "No experiments found" displayed. To the right of the table, it says "0 items selected".

At the bottom of the interface, there are two buttons: "DELETE" and "ADD TO PROJECT".

Step 2 : 실험 생성

The screenshot displays the Microsoft Azure Machine Learning Studio interface. At the top, the header shows 'Microsoft Azure Machine Learning Studio' on the left, and 'time remaining: 07:59', 'Guest', 'SIGN UP', and user icons on the right. Below the header, a blue sidebar on the left contains navigation options: 'PROJECTS', 'NEW', 'DATASET', 'MODULE', 'PROJECT PREVIEW', 'EXPERIMENT', and 'NOTEBOOK PREVIEW'. The main area is titled 'experiments' and includes a search bar labeled 'Search experiment templates'. Below the search bar, a section titled 'Microsoft Samples' displays a grid of experiment templates. A large white arrow with a blue outline points to the 'Blank Experiment' template, which features a plus sign icon. Other templates include 'Sample 1: Download dataset from UCI: Adult 2 class dataset', 'Sample 2: Dataset Processing and Analysis: Auto Imports Regression', 'Sample 3: Cross Validation for Binary Classification: Adult', 'Sample 4: Cross Validation for Regression: Auto Imports Dataset', 'Sample 5: Train, Test, Evaluate for Binary Classification: Adult', and 'Sample 6: Train, Test, Evaluate for Regression: Auto Imports Dataset'. A 'VIEW MORE IN GALLERY' link is visible on the right side of the grid.

Microsoft Azure Machine Learning Studio

time remaining: 07:59 Guest SIGN UP ?

PROJECTS experiments MY EXPERIMENTS SAMPLES

NEW

DATASET MODULE PROJECT PREVIEW EXPERIMENT NOTEBOOK PREVIEW

Search experiment templates

Microsoft Samples

Blank Experiment

Experiment Tutorial

Sample 1: Download dataset from UCI: Adult 2 class dataset

Sample 2: Dataset Processing and Analysis: Auto Imports Regression

Sample 3: Cross Validation for Binary Classification: Adult

Sample 4: Cross Validation for Regression: Auto Imports Dataset

Sample 5: Train, Test, Evaluate for Binary Classification: Adult

Sample 6: Train, Test, Evaluate for Regression: Auto Imports Dataset

VIEW MORE IN GALLERY

Step 3 : 데이터 로딩

- 왼쪽의 Data Input and Output에서 Import data 모듈 드래그
- Data source : Web URL via HTTP
- URL : <http://mlr.cs.umass.edu/ml/machine-learning-databases/iris/iris.data>

Step 3 : 데이터 로딩

The screenshot displays the Microsoft Azure Machine Learning Studio interface. At the top, the header shows 'Microsoft Azure Machine Learning Studio', 'time remaining: 07:55', 'Guest', and a 'SIGN UP' button. The main workspace is titled 'Experiment created on 2016. 11. 29.' and shows a draft experiment with a single step labeled 'Import Data' (1). The left sidebar contains a search bar and a list of experiment items: 'Saved Datasets', 'Data Format Conversions', 'Data Input and Output' (expanded), 'Enter Data Manually', 'Export Data', 'Import Data', 'Unpack Zipped Datasets', 'Data Transformation', 'Feature Selection', 'Machine Learning', 'OpenCV Library Modules', 'Python Language Modules', 'R Language Modules', 'Statistical Functions', 'Text Analytics', and 'Web Service'. The right sidebar shows the 'Properties' tab for the 'Import Data' step, with fields for 'Data source' (Web URL via HTTP), 'Data source URL' (http://mlr.cs.umass.edu/ml/r), 'Data format' (CSV), and checkboxes for 'CSV or TSV has header...' and 'Use cached results'. A 'Quick Help' section at the bottom right provides information on loading data from various sources. The bottom toolbar includes buttons for 'NEW', 'RUN HISTORY', 'SAVE', 'SAVE AS', 'DISCARD CHANGES', 'RUN', 'SET UP WEB SERVICE', and 'PUBLISH TO GALLERY'.

Microsoft Azure Machine Learning Studio

time remaining: 07:55 Guest SIGN UP ?

Experiment created on 2016. 11. 29.

In draft

Draft saved at 오후 5:14:28

Search experiment items

- Saved Datasets
- Data Format Conversions
- Data Input and Output
 - Enter Data Manually
 - Export Data
 - Import Data
 - Unpack Zipped Datasets
- Data Transformation
- Feature Selection
- Machine Learning
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions
- Text Analytics
- Web Service

Import Data 1

Mini Map

Import Data 1

Properties Project

Import Data

Data source: Web URL via HTTP

Data source URL: http://mlr.cs.umass.edu/ml/r

Data format: CSV

☐ CSV or TSV has header...

☐ Use cached results

Quick Help

Load data from sources such as the Web, Azure SQL database, Azure table, Hive table, or Windows Azure BLOB storage. Formerly known as Reader. [\(more help...\)](#)

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

Step 4 : 데이터 전처리

- 왼쪽의 Data Transformation / Manipulation에서 Clean Missing Data, Edit Metadata 모듈 드래그
- Edit Metadata의 Launch column selector를 클릭한 후 All columns, All labels 선택
- New column names : f1,f2,f3,f4,label

Step 4 : 데이터 전처리

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The top navigation bar includes the Microsoft Azure Machine Learning Studio logo, a search bar, and user information (Guest, time remaining: 07:54, SIGN UP button, and icons for help, settings, and profile). The main workspace shows a workflow diagram with three steps: Import Data, Clean Missing Data, and Edit Metadata (labeled with a '1' in a circle). A Mini Map on the left provides a zoomed-out view of the workflow. The left sidebar contains a search bar and a list of experiment items under the 'Data Transformation' category, including Filter, Learning with Counts, Manipulation, Add Columns, Add Rows, Apply SQL Transform..., Clean Missing Data, Convert to Indicator..., Edit Metadata, Group Categorical Va..., Join Data, Remove Duplicate Ro..., Select Columns in Da..., Select Columns Trans..., and SMOTE. The right sidebar shows the 'Properties' panel for the 'Edit Metadata' step, with options for 'Column' (Selected columns: All columns, All labels), 'Data type' (Unchanged), 'Categorical' (Unchanged), 'Fields' (Unchanged), and 'New column names' (f1,f2,f3,f4,label). A 'Quick Help' section at the bottom right provides information about the 'Edit Metadata' step.

Microsoft Azure Machine Learning Studio

time remaining: 07:54 Guest SIGN UP ?

Experiment created on 2016. 11. 29.

In draft

Draft saved at 오후 5:17:11

Search experiment items

Data Transformation

- Filter
- Learning with Counts
- Manipulation
 - Add Columns
 - Add Rows
 - Apply SQL Transform...
 - Clean Missing Data
 - Convert to Indicator ...
 - Edit Metadata
 - Group Categorical Va...
 - Join Data
 - Remove Duplicate Ro...
 - Select Columns in Da...
 - Select Columns Trans...
 - SMOTE
- Sample and Split

Import Data

Clean Missing Data

Edit Metadata 1

Mini Map

Import Data

Clean Missing Data

Edit Metadata 1

Properties Project

Edit Metadata

Column

Selected columns:
All columns
All labels

Launch column selector

Data type

Unchanged

Categorical

Unchanged

Fields

Unchanged

New column names

f1,f2,f3,f4,label

Quick Help

Edits metadata associated with columns in a dataset. Formerly known as Metadata Editor.
(more help...)

+ NEW

RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

Step 5 : Train/Test 나누기

- 왼쪽의 Data Transformation / Sample and Split 에서 Split data 모듈 드래그
- Fraction 0.8으로 설정

Step 5 : Train/Test 나누기

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The top navigation bar includes the logo, "Microsoft Azure Machine Learning Studio", and user information: "time remaining: 07:46", "Guest", and a "SIGN UP" button. The main workspace shows an experiment titled "Experiment created on 2016. 11. 29." in draft status, saved at 5:24:16 PM. The left sidebar contains a search bar and a list of experiment items: Data Format Conversions, Data Input and Output, Data Transformation (selected), Filter, Learning with Counts, Manipulation, Sample and Split (expanded), Partition and Sample, Split Data, Scale and Reduce, Feature Selection, Machine Learning, OpenCV Library Modules, Python Language Modules, R Language Modules, Statistical Functions, and Text Analytics. The central workspace displays a vertical flowchart with four modules: "Import Data", "Clean Missing Data", "Edit Metadata", and "Split Data". The "Split Data" module is highlighted with a blue border and numbered 1 and 2. A "Mini Map" in the bottom-left corner provides a smaller overview of the entire pipeline. The right sidebar shows the "Properties" panel for the "Split Data" module, with settings for "Splitting mode" (Split Rows), "Fraction of rows in the first set" (0.8), "Randomized split" (checked), "Random seed" (0), and "Stratified split" (False). A "Quick Help" section at the bottom right explains the function: "Split the rows of a dataset into two distinct sets (more help...)". The bottom toolbar contains icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

Step 6 : Model 생성

- 왼쪽의 Machine Learning / Train에서 Train Model 모듈 드래그, Machine Learning / Initialize Model / Classification에서 Multiclass Logistic Regression 모듈 드래그 후 연결

Step 6 : Model 생성

- 왼쪽의 Machine Learning / Train에서 Train Model 모듈 드래그, Machine Learning / Initialize Model / Classification에서 Multiclass Logistic Regression 모듈 드래그 후 연결
- Train Model의 launch column selector를 누르고 column names에 label 입력

Step 6 : Model 생성

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The top navigation bar includes the title "Microsoft Azure Machine Learning Studio", a timer "time remaining: 07:43", a "Guest" user profile, and a "SIGN UP" button. The left sidebar contains a search bar and a list of experiment items: Data Transformation, Feature Selection, Machine Learning (selected), Evaluate, Initialize Model, Anomaly Detection, Classification, Clustering, Regression, Score, Train, OpenCV Library Modules, Python Language Modules, R Language Modules, Statistical Functions, Text Analytics, Web Service, and Deprecated.

The main workspace shows a workflow diagram titled "Experiment created on 2016. 11. 29." with the status "In draft" and "Draft saved at 오후 5:27:39". The workflow consists of the following steps: "Import Data" → "Clean Missing Data" → "Edit Metadata" → "Split Data". From "Split Data", the flow branches into "Multiclass Logistic Regression" and "Train Model" (labeled with a circled '1'). A "Mini Map" in the bottom-left corner provides a zoomed-out view of the entire workflow. The bottom toolbar contains icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

On the right side, the "Properties" panel is active for the "Train Model" step. It shows the "Label column" property with "Selected columns: Column names: label" and a "Launch column selector" button. Below this, the "Quick Help" section provides instructions: "Train a previously created classification or regression model (more help...)"

Step 7 : Model 평가

- 왼쪽의 Machine Learning / Score 에서 Score Model 모듈 드래그, Machine Learning / Evaluate 에서 Evaluate Model 모듈 드래그 후 연결
- Train Model의 launch column selector를 누르고 column names에 label 입력

Step 7 : Model 평가

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The top header shows the title "Microsoft Azure Machine Learning Studio", a timer "time remaining: 07:40", the user "Guest", and a "SIGN UP" button. The main workspace is titled "Experiment created on 2016. 11. 29." and shows a workflow diagram. The workflow starts with "Import Data", followed by "Clean Missing Data", "Edit Metadata", and "Split Data". From "Split Data", the flow branches into "Multiclass Logistic Regression" and "Train Model". Both lead to "Score Model", which then leads to "Evaluate Model". A "Mini Map" in the bottom left corner provides a overview of the workflow. The left sidebar contains a navigation pane with categories: "Data Transformation", "Feature Selection", "Machine Learning", "Evaluate", "Initialize Model", "Score", and "Train". The "Evaluate" section is expanded, showing "Cross Validate Model", "Evaluate Model", and "Evaluate Recommend...". The right sidebar contains "Properties" and "Project" tabs, with "Experiment Properties" selected. It includes a "Summary" section for entering a description and a "Description" section for detailed notes. The bottom status bar includes icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

Microsoft Azure Machine Learning Studio

time remaining: 07:40 Guest SIGN UP

Experiment created on 2016. 11. 29.

In draft

Draft saved at 오후 5:30:17

Search experiment items

Machine Learning

- Evaluate
 - Cross Validate Model
 - Evaluate Model
 - Evaluate Recommend...
- Initialize Model
 - Anomaly Detection
 - Classification
 - Clustering
 - Regression
- Score
 - Apply Transformation
 - Assign Data to Clusters
 - Score Matchbox Rec...
 - Score Model
- Train

Import Data

Clean Missing Data

Edit Metadata

Split Data

Multiclass Logistic Regression

Train Model

Score Model

Evaluate Model

Mini Map

Properties Project

Experiment Properties

STATUS CODE InDraft

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Description

Enter the detailed description for your experiment.

Quick Help

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

Step 8 : 실행

- 오른쪽 하단의 RUN 클릭
- 작업이 완료되면 Evaluate Model을 오른쪽 클릭, Evaluation results/Visualize 옵션 클릭

Step 8 : 실행

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The top navigation bar includes the Microsoft Azure Machine Learning Studio logo, a search bar, and user information (time remaining: 07:38, Guest, SIGN UP, and user icons). The left sidebar contains a search bar and a list of experiment items: Saved Datasets, Data Format Conversions, Data Input and Output, Data Transformation, Feature Selection, and Machine Learning. The Machine Learning section is expanded, showing Evaluate, Initialize Model, and Score. The Evaluate section includes Cross Validate Model, Evaluate Model, and Evaluate Recommendation... The Initialize Model section includes Anomaly Detection, Classification, Clustering, and Regression. The Score section includes Apply Transformation and Assign Data to Clusters.

The main workspace shows a workflow diagram titled "Experiment created on 2016. 11. 29." The workflow consists of the following steps: Import Data, Clean Missing Data, Edit Metadata, Split Data, Multiclass Logistic Regression, Train Model, Score Model, and Evaluate Model. The Evaluate Model step is highlighted with a blue box and a circled number 1. A context menu is open over the Evaluate Model step, showing options: Download, Save as Dataset, Save as Trained Model, Save as Transform, Visualize, Generate Data Access Code..., Open in a new Notebook, Evaluation results, View Log, Edit Comment, Run selected, and Help.

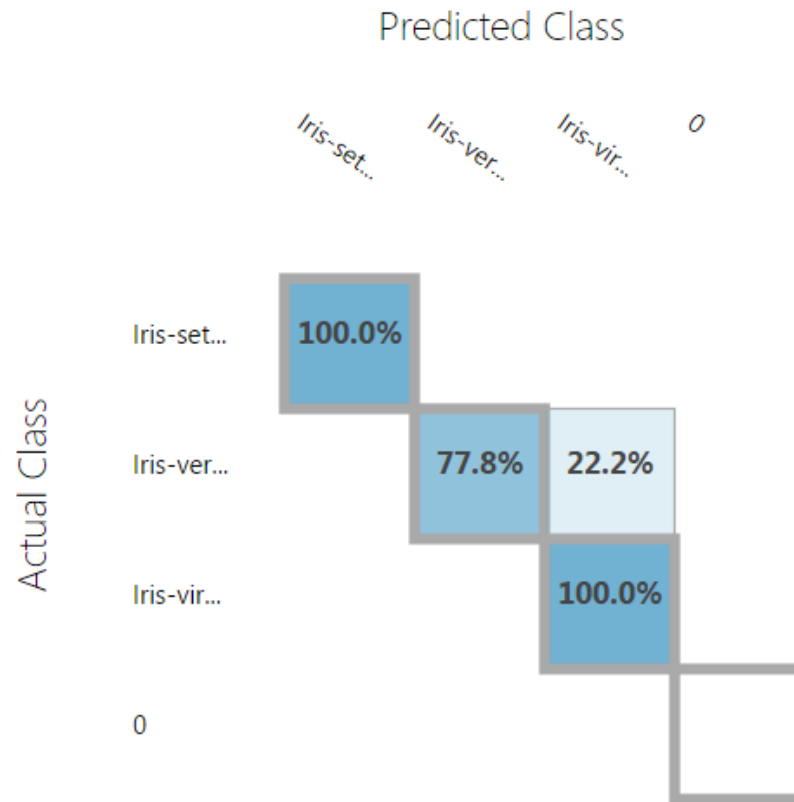
The right sidebar displays the "Properties" tab for the "Evaluate Model" step. It shows the following details:

- START TIME: 11/29/2016...
- END TIME: 11/29/2016...
- ELAPSED TIME: 0:00:02.750
- STATUS CODE: Finished
- STATUS DETAILS: None

A "View output log" link is also present. The bottom status bar includes a "NEW" button, a "RUN HISTORY" button, and "SAVE", "SAVE AS", "DISCARD CHANGES", and "RUN" buttons.

Step 8 : 실행

Confusion Matrix



감사합니다
