**Introduction to AI**

**QUESTION**

**1. You want to create a model to predict sales of ice cream based on historic data that includes daily ice cream sales totals and weather measurements. Which Azure service should you use?**

Azure Machine Learning

That is correct. Azure Machine Learning enables you to train a predictive model from the existing data.

QnA Maker

Text Analytics

**2. You want to train a model that classifies images of dogs and cats based on a collection of your own digital photographs. Which Azure service should you use?**

Azure Bot Service

Custom Vision

That is correct. Custom Vision enables you to train an image classification model based on your own images.

Computer Vision

**3. You are designing an AI application that uses computer vision to detect cracks in car windshields, and warns drivers when a windshield should be repaired or replaced. When tested in good lighting conditions, the application successfully detects 99% of dangerously damaged glass. Which of the following statements should you include in the application's user interface?**

When used in good lighting conditions, this application can be used to identify potentially dangerous cracks and defects in windshields. If you suspect your windshield is damaged, even if the application does not detect any defects, you should have it inspected by a professional.

That is correct. You should be transparent about the limitations of the application.

This application detects damage in your windshield. If the application detects a defect, have the windshield replaced or repaired. If no defect is detected, you're good to go!

This application detects damage in any glass surface, but you must accept responsibility for using it only in appropriate lighting conditions.

**Module 2:**

**Create no-code predictive models with Azure Machine Learning**

**Chapter 1:**

**Use automated machine learning in Azure Machine Learning**

**1. An automobile dealership wants to use historic car sales data to train a machine learning model. The model should predict the price of a pre-owned car based on its make, model, engine size, and mileage. What kind of machine learning model should the dealership use automated machine learning to create?**

**Classification**

**Regression**

**That is correct. To predict a numeric value, use a regression model.**

**Time series forecasting**

**2. A bank wants to use historic loan repayment records to categorize loan applications as low-risk or high-risk based on characteristics like the loan amount, the income of the borrower, and the loan period. What kind of machine learning model should the bank use automated machine learning to create?**

**Classification**

**That is correct. To predict a category, or class, use a classification model.**

**Regression**

**Time series forecasting**

**3. You want to use automated machine learning to train a regression model with the best possible R2 score. How should you configure the automated machine learning experiment?**

**Set the Primary metric to R2 score**

**That is correct. The primary metric determines the metric used to evaluate the best performing model.**

**Block all algorithms other than GradientBoosting**

**Enable featurization**

**Chapter 2 :**

**Create a Regression Model with Azure Machine Learning designer**

**1. You are creating a training pipeline for a regression model, using a dataset that has multiple numeric columns in which the values are on different scales. You want to transform the numeric columns so that the values are all on a similar scale based relative to the minimum and maximum values in each column. Which module should you add to the pipeline?**

**Select Columns in a Dataset**

**Normalize Data**

**That is correct. When you need to transform numeric data to be on a similar scale, use a Normalize Data module.**

**Clean Missing Data**

**2. You use Azure Machine Learning designer to create a training pipeline and an inference pipeline for a regression model. Now you plan to deploy the inference pipeline as a real-time service. What kind of compute target should you create to host the service?**

**Compute Instance**

**Compute Cluster**

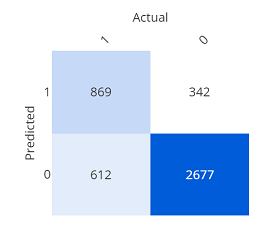
**Inference Cluster**

**That is correct. Use an Inference Cluster for deployment.**

* **Mean Absolute Error (MAE)**: The average difference between predicted values and true values. This value is based on the same units as the label, in this case dollars. The lower this value is, the better the model is predicting.
* **Root Mean Squared Error (RMSE)**: The square root of the mean squared difference between predicted and true values. The result is a metric based on the same unit as the label (dollars). When compared to the MAE (above), a larger difference indicates greater variance in the individual errors (for example, with some errors being very small, while others are large).
* **Relative Squared Error (RSE)**: A relative metric between 0 and 1 based on the square of the differences between predicted and true values. The closer to 0 this metric is, the better the model is performing. Because this metric is relative, it can be used to compare models where the labels are in different units.
* **Relative Absolute Error (RAE)**: A relative metric between 0 and 1 based on the absolute differences between predicted and true values. The closer to 0 this metric is, the better the model is performing. Like RSE, this metric can be used to compare models where the labels are in different units.
* **Coefficient of Determination (R2)**: This metric is more commonly referred to as *R-Squared*, and summarizes how much of the variance between predicted and true values is explained by the model. The closer to 1 this value is, the better the model is performing.

**Chapter 3**

**Create a classification model with Azure Machine Learning designer**



**The confusion matrix shows cases where both the predicted and actual values were 1 (known as true positives) at the top left,**

**and cases where both the predicted and the actual values were 0 (true negatives) at the bottom right.**

**The other cells show cases where the predicted and actual values differ (false positives and false negatives).**

**The cells in the matrix are colored so that the more cases represented in the cell, the more intense the color - with the result that you can identify a model that predicts accurately for all classes by looking for a diagonal line of intensely colored cells from the top left to the bottom right (in other words, the cells where the predicted values match the actual values). For a multi-class classification model (where there are more than two possible classes), the same approach is used to tabulate each possible combination of actual and predicted value**

**counts - so a model with three possible classes would result in a 3x3 matrix with a diagonal line of cells where the predicted and actual labels match.**

Review the metrics to the left of the confusion matrix, which include:

* **Accuracy**: The ratio of correct predictions (true positives + true negatives) to the total number of predictions. In other words, what proportion of diabetes predictions did the model get right?
* **Precision**: The fraction of positive cases correctly identified (the number of true positives divided by the number of true positives plus false positives). In other words, out of all the patients that the model predicted as having diabetes, how many are actually diabetic?
* **Recall**: The fraction of the cases classified as positive that are actually positive (the number of true positives divided by the number of true positives plus false negatives). In other words, out of all the patients who actually have diabetes, how many did the model identify?
* **F1 Score**: An overall metric that essentially combines precision and recall.
* *We'll return to****AUC****later*.

Of these metric, accuracy is the most intuitive. However, you need to be careful about using simple accuracy as a measurement of how well a model works. Suppose that only 3% of the population is diabetic. You could create a model that always predicts 0 and it would be 97% accurate - just not very useful! For this reason, most data scientists use other metrics like precision and recall to assess classification model performance.

Above the list of metrics, note that there's a **Threshold** slider. Remember that what a classification model predicts is the probability for each possible class. In the case of this binary classification model, the predicted probability for a *positive* (that is, diabetic) prediction is a value between 0 and 1. By default, a predicted probability for diabetes above 0.5 results in a class prediction of 1, while a prediction below this threshold means that there's a greater probability of the patient **not** having diabetes (remember that the probabilities for all classes add up to 1), so the predicted class would be 0. Try moving the threshold slider and observe the effect on the confusion matrix. If you move it all the way to the left (0), the Recall metric becomes 1, and if you move it all the way to the right (1), the Recall metric becomes 0.

Look above the Threshold slider at the ROC curve (ROC stands for received operator characteristic, but most data scientists just call it a ROC curve). Another term for recall is True positive rate, and it has a corresponding metric named False positive rate, which measures the number of negative cases incorrectly identified as positive compared the number of actual negative cases. Plotting these metrics against each other for every possible threshold value between 0 and 1 results in a curve. In an ideal model, the curve would go all the way up the left side and across the top, so that it covers the full area of the chart. The larger the area under the curve (which can be any value from 0 to 1), the better the model is performing - this is the AUC metric listed with the other metrics below. To get an idea of how this area represents the performance of the model, imagine a straight diagonal line from the bottom left to the top right of the ROC chart. This represents the expected performance if you just guessed or flipped a coin for each patient - you could expect to get around half of them right, and half of them wrong, so the area under the diagonal line represents an AUC of 0.5. If the AUC for your model is higher than this for a binary classification model, then the model performs better than a random guess.

1. You are using Azure Machine Learning designer to create a training pipeline for a binary classification model. You have added a dataset containing features and labels, a Two-Class Decision Forest module, and a Train Model module. You plan to use Score Model and Evaluate Model modules to test the trained model with a subset of the dataset that was not used for training. Which additional kind of module should you add?

Join Data

Split Data

That is correct. Use a Split Data module to randomly split a dataset into training and validation subsets.

Select Columns in Dataset

2. You use an Azure Machine Learning designer pipeline to train and test a binary classification model. You review the model's performance metrics in an Evaluate Model module, and note that it has an AUC score of 0.3. What can you conclude about the model?

The model can explain 30% of the variance between true and predicted labels.

The model predicts accurately for 70% of test cases.

The model performs worse than random guessing.

That is correct. An AUC of 0.5 is what you'd expect with random prediction of a binary model.

3. You use Azure Machine Learning designer to create a training pipeline for a classification model. What must you do before deploying the model as a service?

Create an inference pipeline from the training pipeline

That is correct. You must create an inference pipeline to deploy as a service.

Add an Evaluate Model module to the training pipeline

Clone the training pipeline with a different name

**Chapter 4**

**Create a Clustering Model with Azure Machine Learning designer**

Clustering is a form of machine learning that is used to group similar items into clusters based on their features. For example, a researcher might take measurements of penguins, and group them based on similarities in their proportions.

Clustering is an example of unsupervised machine learning, in which you train a model to separate items into clusters based purely on their characteristics, or features. There is no previously known cluster value (or label) from which to train the model.

* **Average Distance to Other Center**: This indicates how close, on average, each point in the cluster is to the centroids of all other clusters.
* **Average Distance to Cluster Center**: This indicates how close, on average, each point in the cluster is to the centroid of the cluster.
* **Number of Points**: The number of points assigned to the cluster.
* **Maximal Distance to Cluster Center**: The maximum of the distances between each point and the centroid of that point’s cluster. If this number is high, the cluster may be widely dispersed. This statistic in combination with the **Average Distance to Cluster Center** helps you determine the cluster’s *spread*.

**1. You are using an Azure Machine Learning designer pipeline to train and test a K-Means clustering model. You want your model to assign items to one of three clusters. Which configuration property of the K-Means Clustering module should you set to accomplish this?**

**Set Number of Centroids to 3**

**That is correct. To create K clusters, you must set the number of centroids to K**

**Set Random number seed to 3**

**Set Iterations to 3**

**2. You use Azure Machine Learning designer to create a training pipeline for a clustering model. Now you want to use the model in an inference pipeline. Which module should you use to infer cluster predictions from the model?**

**Score Model**

**Assign Data to Clusters**

**That is correct. use the Assign Data to Clusters module to generate cluster predictions from a trained clustering model**

**Train Clustering Model**

**Module 3**

**Explore computer vision in Microsoft Azure**

**Chapter 1**

**Analysis Image with computer vision service**

*Computer vision* is one of the core areas of artificial intelligence (AI), and focuses on creating solutions that enable AI-enabled applications to "see" the world and make sense of it.

Of course, computers don't have biological eyes that work the way ours do, but they are capable of processing images; either from a live camera feed or from digital photographs or videos. This ability to process images is the key to creating software that can emulate human visual perception.

To an AI application, an image is just an array of pixel values. These numeric values can be used as *features* to train machine learning models that make predictions about the image and its contents.

In Microsoft Azure, the **Computer Vision** cognitive service uses pre-trained models to analyze images, enabling software developers to easily build applications that can:

* Interpret an image and suggest an appropriate caption.
* Suggest relevant *tags* that could be used to index an image.
* Categorize an image.
* Identify objects in an image.
* Detect faces and people in an image.
* Recognize celebrities and landmarks in an image.
* Read text in an image.

## Azure resources for Computer Vision

To use the Computer Vision service, you need to create a resource for it in your Azure subscription. You can use either of the following resource types:

* **Computer Vision**: A specific resource for the Computer Vision service. Use this resource type if you don't intend to use any other cognitive services, or if you want to track utilization and costs for your Computer Vision resource separately.
* **Cognitive Services**: A general cognitive services resource that includes Computer Vision along with many other cognitive services; such as Text Analytics, Translator Text, and others. Use this resource type if you plan to use multiple cognitive services and want to simplify administration and development.

**1. You want to use the Computer Vision service to analyze images. You also want to use the Text Analytics service to analyze text. You want developers to require only one key and endpoint to access all of your services. What kind of resource should you create in your Azure subscription?**

**Computer Vision**

**Cognitive Services**

**Correct. A Cognitive Services resource support both Computer Vision and Text Analytics.**

**Custom Vision**

**2. You want to use the Computer Vision service to identify the location of individual items in an image. Which of the following features should you retrieve?**

**Objects**

**Correct. Computer Vision returns objects with a bounding box to indicate their location in the image.**

**Tags**

**Categories**

**3. You want to use the Computer Vision service to analyze images of locations and identify well-known buildings. What should you do?**

**Retrieve the objects in the image.**

**Retrieve the categories for the image, specifying the celebrities domain**

**Retrieve the categories for the image, specifying the landmarks domain**

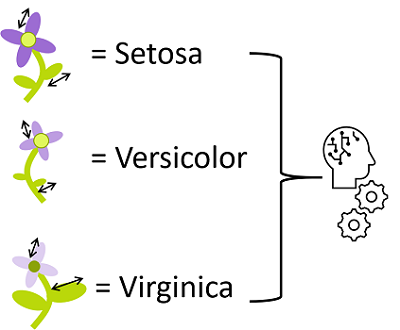
**Correct. The landmarks domain includes many well-known buildings around the world.**

**Chapter 2**

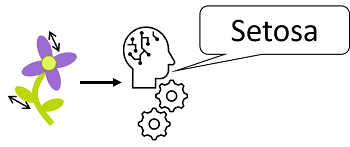
**Classify images with the Custom Vision service**

Classification is a machine learning technique that you can use to predict which category, or class, something belongs to. Classification machine learning models use a set of inputs, which we call features, to calculate a probability score for each possible class and predict a label that indicates the most likely class that an object belongs to.

For example, the features of a flower might include the measurements of its petals, stem, sepals, and other quantifiable characteristics. A machine learning model could be trained by applying an algorithm to these measurements that calculates the most likely species of the flower - its class.



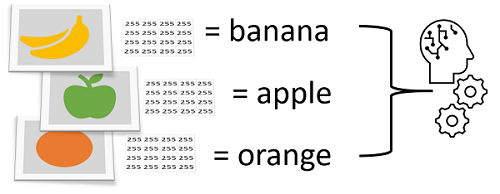
To create a classification model, you need data that consists of features for which the labels are already known. You use this to train the model so that it determines a relationship between the features and the corresponding label. Then, after the model has been trained, you can use it with new sets of features to predict unknown label values.



**Image Classification**

Image classification is a machine learning technique in which the object being classified is an image, such as a photograph.

As with any form of classification, creating an image classification solution involves training a machine learning model using a set of existing data for which the class is already known. In this case, the existing data consists of a set of categorized images. Digital images are made up of an array of pixel values, and these are used as features to train the model based on the known image classes.



Most modern image classification solutions are based on deep learning techniques that make use of convolutional neural networks (CNNs) to uncover patterns in the pixels that correspond to particular classes. Training an effective CNN is a complex task that requires considerable expertise in data science and machine learning. However, common techniques used to train image classification models have been encapsulated into the **Custom Vision** cognitive service in Microsoft Azure; making it easy to train a model and publish it as a software service with minimal knowledge of deep learning techniques.

## Uses of image classification

Some potential uses for image classification might include:

* Product identification - performing visual searches for specific products in online searches or even, in-store using a mobile device.
* Disaster investigation - evaluating key infrastructure for major disaster preparation efforts. For example, aerial surveillance images may show bridges and classify them as such. Anything classified as a bridge could then be marked for emergency preparation and investigation.
* Medical diagnosis - evaluating images from X-ray or MRI devices could quickly classify specific issues found as cancerous tumors, or many other medical conditions related to medical imaging diagnosis.

Creating an image classification solution with Custom Vision consists of two main tasks. First you must use existing images to train the model, and then you must publish the model so that client applications can use it to generate predictions.

For each of these tasks, you need a resource in your Azure subscription. You can use the following types of resource:

* **Custom Vision**: A dedicated resource for the custom vision service, which can be *training*, a *prediction*, or *both* resources.
* **Cognitive Services**: A general cognitive services resource that includes Custom Vision along with many other cognitive services. You can use this type of resource for *training*, *prediction*, or both.

## Model evaluation

Model training process is an iterative process in which the Custom Vision service repeatedly trains the model using some of the data, but holds some back to evaluate the model. At the end of the training process, the performance for the trained model is indicated by the following evaluation metrics:

* **Precision**: What percentage of the class predictions made by the model were correct? For example, if the model predicted that 10 images are oranges, of which eight were actually oranges, then the precision is 0.8 (80%).
* **Recall**: What percentage of class predictions did the model correctly identify? For example, if there are 10 images of apples, and the model found 7 of them, then the recall is 0.7 (70%).
* **Average Precision (AP)**: An overall metric that takes into account both precision and recall).

**1. You plan to use the Custom Vision service to train an image classification model. You want to create a resource that can only be used for model training, and not for prediction. Which kind of resource should you create in your Azure subscription?**

**Custom Vision**

**Correct: When you create a Custom Vision resource, you can specify whether it is to be used for training, prediction, or both.**

**Cognitive Services**

**Computer Vision**

**2. You train an image classification model that achieves less than satisfactory evaluation metrics. How might you improve it?**

**Reduce the size of the images used to train the model.**

**Add a new label for "unknown" classes.**

**Add more images to the training set.**

**Correct: Generally, adding more images to the project an retraining the model is likely to improve performance.**

**3. You have published an image classification model. What information must you provide to developers who want to use it?**

**Only the project ID.**

**The project ID, the model name, and the key and endpoint for the prediction resource**

**Correct: To use a published model, you need the project ID, the model name, and the key and endpoint for the prediction resource.**

**The project ID, iteration number, and the key and endpoint for the training resource.**