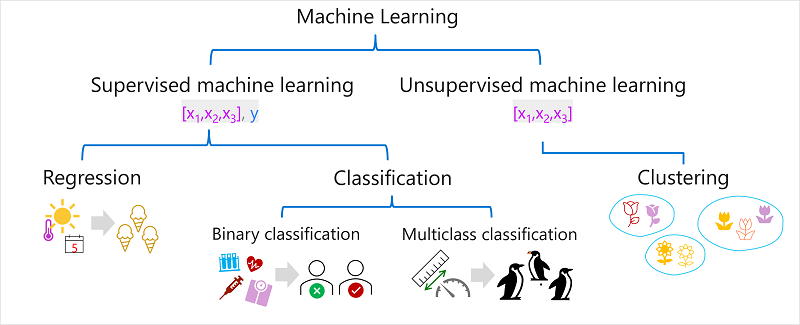
**What is machine learning?**

Machine learning has its origins in statistics and mathematical modeling of data. The fundamental idea of machine learning is to use data from past observations to predict unknown outcomes or values.

For example:

* The proprietor of an ice cream store might use an app that combines historical sales and weather records to predict how many ice creams they're likely to sell on a given day, based on the weather forecast.



Supervised machine learning is a type of machine learning that uses labeled data to train algorithms to recognize patterns and predict outcomes.

### **Regression**

Regression is a form of **supervised machine learning** in which the **label predicted by the model is a numeric value**. For example:

* The number of ice creams sold on a given day, based on the temperature, rainfall, and windspeed.
* The selling price of a property based on its size in square feet, the number of bedrooms it contains, and socio-economic metrics for its location.

### **Classification**

Classification is a form of **supervised machine learning** in which the **label represents a categorization, or class.** There are two common classification scenarios.

#### **Binary classification**

Binary classification is a machine learning algorithm that categorizes data into one of two classes, with the results being either positive or negative.

For example:

* Whether a patient is at risk for diabetes based on clinical metrics like weight, age, blood glucose level, and so on.
* Whether a bank customer will default on a loan based on income, credit history, age, and other factors.

**Multiclass classification**

Multiclass classification is the process of assigning entities with more than two classes.

For example,

* The species of a penguin (Adelie, Gentoo, or Chinstrap) based on its physical measurements.
* The genre of a movie (comedy, horror, romance, adventure, or science fiction) based on its cast, director, and budget.

## **Unsupervised machine learning**

Unsupervised machine learning involves training models using data that consists only of feature values without any known labels. Unsupervised machine learning algorithms determine relationships between the features of the observations in the training data.

### **Clustering**

The most common form of **unsupervised machine learning** is clustering**. A clustering algorithm identifies similarities between observations based on their features, and groups them into discrete clusters**.

For example:

* Group similar flowers based on their size, number of leaves, and number of petals.
* Identify groups of similar customers based on demographic attributes and purchasing behavior.

# Regression

**Regression models are trained to predict numeric label values based on training data that includes both features and known labels.** The process for training a regression model (or indeed, any supervised machine learning model) involves multiple iterations in which you use an appropriate algorithm (usually with some parameterized settings) to train a model, evaluate the model's predictive performance, and refine the model by repeating the training process with different algorithms and parameters until you achieve an acceptable level of predictive accuracy.

### **Regression evaluation metrics**

#### **Mean Absolute Error (MAE)**

which calculates the **absolute difference between actual and predicted values.**

To better understand, let’s take an example you have input data and output data and use Linear Regression, which draws a best-fit line.

Now you have to find the MAE of your model which is basically a mistake made by the model known as an error.

Now find the difference between the actual value and predicted value that is an absolute error but we have to find the mean absolute of the complete dataset. A diagram of mathematical equations

Description automatically generated

**OR**

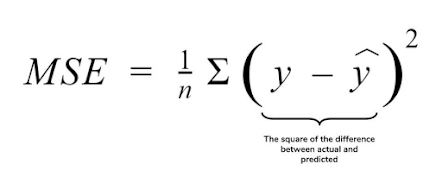
The variance in this example indicates by how many ice creams each prediction was wrong. It doesn't matter if the prediction was over or under the actual value (so for example, -3 and +3 both indicate a variance of 3). This metric is known as the absolute error for each prediction, and can be summarized for the whole validation set as the mean absolute error (MAE).

In the ice cream example, the mean (average) of the absolute errors (2, 3, 3, 1, 2, and 3) is 2.3

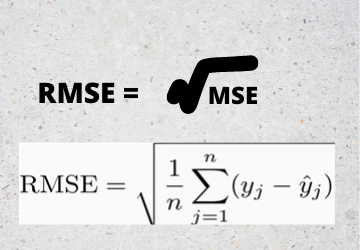
#### **Mean Squared Error (MSE)**

Mean squared error states **that finding the squared difference between actual and predicted value.**

we perform squared to **avoid the cancellation of negative terms and it is the benefit of MSE.**



#### **Root Mean Squared Error (RMSE)**

A mathematical equation with numbers and symbols

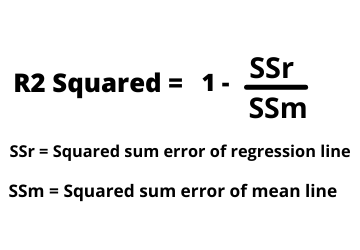
Description automatically generated with medium confidence

#### **Coefficient of determination (R2)**

R2 score is a metric that tells the performance of your model.

The calculation for R2 is more complex than for the previous metrics. It compares the sum of squared differences between predicted and actual labels with the sum of squared differences between the actual label values and the mean of actual label values, like this:

R2 = 1- ∑(y-ŷ)2 ÷ ∑(y-ȳ)2



Don't worry too much if that looks complicated; most machine learning tools can calculate the metric for you. The important point is that the result is a value between 0 and 1 that describes the proportion of variance explained by the model. In simple terms, the closer to 1 this value is, the better the model is fitting the validation data. In the case of the ice cream regression model, the R2 calculated from the validation data is 0.95.

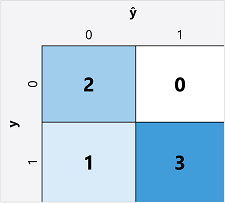
# Binary classification

Classification, like regression, is a supervised machine learning technique; and therefore follows the same iterative process of training, validating, and evaluating models. Instead of calculating numeric values like a regression model, the algorithms used to train classification models calculate probability values for class assignment and the evaluation metrics used to assess model performance compare the predicted classes to the actual classes.

Binary classification algorithms are used to train a model that predicts one of two possible labels for a single class. Essentially, predicting true or false. In most real scenarios, the data observations used to train and validate the model **consist of multiple feature (x) values and a y value that is either 1 or 0.**

### **Binary classification evaluation metrics**

the first step in calculating evaluation metrics for a binary classification model is usually to create a matrix of the number of correct and incorrect predictions for each possible class label:



This visualization is called a confusion matrix, and it shows the prediction totals where:

* ŷ=0 and y=0: True negatives (TN)
* ŷ=1 and y=0: False positives (FP)
* ŷ=0 and y=1: False negatives (FN)
* ŷ=1 and y=1: True positives (TP)

#### **Accuracy**

The simplest metric you can calculate from the confusion matrix is accuracy - the proportion of predictions that the model got right. Accuracy is calculated as:

(TN+TP) ÷ (TN+FN+FP+TP)

In the case of our diabetes example, the calculation is:

(2+3) ÷ (2+1+0+3)

= 5 ÷ 6

= 0.83

#### **Recall**

Recall is a metric that measures the proportion of positive cases that the model identified correctly. In other words, compared to the number of patients who have diabetes, how many did the model predict to have diabetes?

The formula for recall is:

TP ÷ (TP+FN)

For our diabetes example:

3 ÷ (3+1)

= 3 ÷ 4

= 0.75

So our model correctly identified 75% of patients who have diabetes as having diabetes.

#### **Precision**

Precision is a similar metric to recall, but measures the proportion of predicted positive cases where the true label is actually positive. In other words, what proportion of the patients predicted by the model to have diabetes actually have diabetes?

The formula for precision is:

TP ÷ (TP+FP)

For our diabetes example:

3 ÷ (3+0)

= 3 ÷ 3

= 1.0

So 100% of the patients predicted by our model to have diabetes do in fact have diabetes.

#### **F1-score**

F1-score is an overall metric that combined recall and precision. The formula for F1-score is:

(2 x Precision x Recall) ÷ (Precision + Recall)

For our diabetes example:

(2 x 1.0 x 0.75) ÷ (1.0 + 0.75)

= 1.5 ÷ 1.75

= 0.86

#### **Area Under the Curve (AUC)**

Another name for recall is the true positive rate (TPR), and there's an equivalent metric called the false positive rate (FPR) that is calculated as FP÷(FP+TN). We already know that the TPR for our model when using a threshold of 0.5 is 0.75, and we can use the formula for FPR to calculate a value of 0÷2 = 0.

# **Deep learning**

Deep learning is an advanced form of machine learning that tries to emulate the way the human brain learns. The key to deep learning is the creation of an artificial neural network that simulates electrochemical activity in biological neurons by using mathematical functions,

# Azure Machine Learning

Microsoft Azure Machine Learning is a cloud service for training, deploying, and managing machine learning models. It's designed to be used by data scientists, software engineers, devops professionals, and others to manage the end-to-end lifecycle of machine learning projects, including:

* Exploring data and preparing it for modeling.
* Training and evaluating machine learning models.
* Registering and managing trained models.
* Deploying trained models for use by applications and services.
* Reviewing and applying responsible AI principles and practices.