CoGrammar

Introduction to Machine Learning

SKILLS FOR LIFE

SKILLS BOOTCAMPS

Department for Education

Data Science Lecture Housekeeping

- The use of disrespectful language is prohibited in the questions, this is a supportive, learning environment for all - please engage accordingly. **(FBV: Mutual Respect.)**
- No question is daft or silly - **ask them!**
- There are **Q&A sessions** midway and at the end of the session, should you wish to ask any follow-up questions. Moderators are going to be answering questions as the session progresses as well.
- If you have any questions outside of this lecture, or that are not answered during this lecture, please do submit these for upcoming Open Classes. You can submit these questions here: **Open Class Questions**

Data Science Lecture Housekeeping cont.

- For all **non-academic questions**, please submit a query: **www.hyperiondev.com/support**
- Report a **safeguarding** incident: **www.hyperiondev.com/safeguardreporting**
- We would love your **feedback** on lectures: **Feedback on Lectures**

Lecture Objectives

- **Introduction to concepts in Machine Learning, providing a general overview.**
- **Preparing to learn about supervised and unsupervised learning algorithms.**

History of Machine Learning

- ★ Machine learning programs are working behind the scenes to produce and curate our playlists, news feeds, weather reports and email inboxes, they help us find restaurants, translate documents, and even meet potential dates.

- ★ From a business perspective, machine learning-based software is becoming central to many industries, generating demand for experts.

History of Machine Learning

- In 1959, Arthur Samuel, a pioneer in artificial intelligence and gaming defined machine learning as the “field of study that gives computers the ability to learn without explicitly programmed.”

- He is known for developing a program capable of playing checkers. Samuel never programmed exactly which strategies the systems could use. Instead, he devised a way in which the program could learn such strategies through the experience of playing thousands of games.

Defining Machine Learning

- A widely used definition of machine learning is that provided by Tom Mitchell, author of a well-known textbook simply titled, “Machine Learning”. This definition is as follows:
  - “A program can be said to learn from experience(E) with respect to some class of tasks(T) and performance measure(P), if its performance at tasks in T, as measured by P, improves with experience E.”

Defining Machine Learning

- At its essence, machine learning (ML) can be defined as a computational methodology focused on deriving insights from data.
- It enables computers to acquire knowledge from past observations and independently make predictions or decisions without relying on explicit programming instructions.
- By leveraging data-driven patterns and algorithms, machine learning enables automated systems to adapt and improve their performance over time.

Defining Machine Learning

The term “machine learning” is often used interchangeably with the term “artificial intelligence” (AI). While the two are very much related, they are not the same thing. A simple way to look at it for our purposes is to see Machine Learning as a type of artificial intelligence.

Any program that completes a task in a way that can be considered human-like can be considered an example of artificial intelligence, but only tasks that solve the task by learning without preprogramming are machine learning programs.

Input and Output

- Whatever it is that we want a machine learning algorithm to learn, we first need to express it numerically. The machine-readable version of a task consists of an input and an output.

- The input is whatever we want the algorithm to learn from, and the output is the outcome we want the algorithm to be able to produce.

Supervised Learning

- In supervised learning problems, a program predicts an output given an input by learning from pairs of inputs and outputs (labels); that is, the program learns from examples that have had the right answers assigned to them beforehand.
- These assignments are often called annotations. Because they are considered the correct answers, they are also called gold labels, gold data or the gold standard. The collection of examples that comprise the supervised experience is called a training set.

Supervised Learning

The collection of data examples used to assess a program's performance is called a test set. Like a student learning in a language course that teaches only through exposure, supervised learning problems see a collection of correct answers to new but similar questions.

Regression

- Regression is a prediction task where a program learns to estimate and predict a continuous output values. It does this by analysing pairs of input features and their corresponding outputs to a training set.

- By analysing the training examples, the program tries to identify patterns and associations that allow it to make precise estimations.

Let’s Breathe!

Regression

- The main objective of regression is to understand the relationship between the input variables and a continuous target variable, enabling the program to make accurate predictions for new inputs that are similar to the training data.
- Commonly used metrics to assess the accuracy of a regression model, which allow for comparing different models or evaluation the performance of a single model
  - R-squared (R2)
  - Mean squared error (MSE)
  - Root mean squared error (RMSE)
  - Mean absolute error (MAE)
  - Mean absolute percentage error (MAPE)

Regression

R2 is known as the coefficient of determination, quantifying the proportion of variance in the target variable that can be explained by the features in the model. It ranges from 0 to 1, with higher values indicating a better fit.

MSE measures the average squared difference between predicted and actual values, providing an overall measure of prediction accuracy.

RMSE is the square root of MSE and represents the the average magnitude of of prediction errors.

Regression

Another metric that provides a measure of the average magnitude of errors is MAE which calculator the average difference between predicted and actual values.

MAPE measures the average percentage difference between predicted and actual values, which is partially useful when the magnitude of errors needs to be access relative to the actual values.

By default, lower values of MSE, RMSE, MAE and MAPE indicate better model performance.

Classification

- Unlike regression tasks, a classification process assumes a program is trained to categorise input data into predefined classes or categories.

- By analysing labelled examples, where each example i already assigned to a specific class, the program learns patterns and relationships between input features and classes. This knowledge allows the program to accurately classify new, unseen data, ensuring they are correctly assigned to their respective classes.

Classification

- The ultimate goal of classification is to develop a model that can make reliable predictions for unknown instances, effectively categorising them into the appropriate classes.
- To evaluate the effectiveness and performance of classification models, several commonly used evaluation metrics are available. Like precision and the F1 score.
- Precision calculates the proportion of of true positive predictions among all positive predictions, indicating the model's ability to minimise false positives.

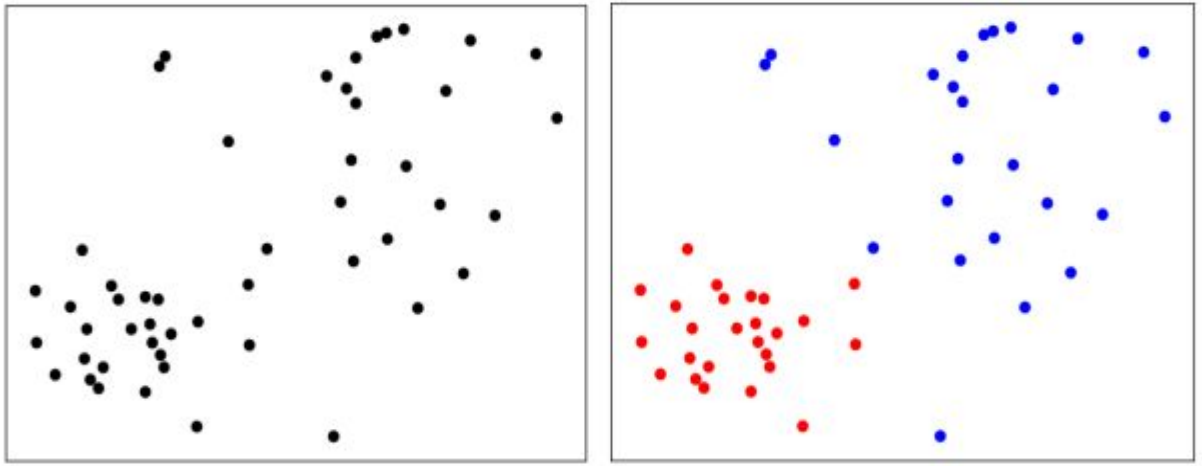
Classification

The F1 score, as the harmonic mean of precision and recall, assesses the model's performance by considering both metrics. Together the discussed metrics provide a comprehensive evaluation of classification models.

Unsupervised Learning

- In unsupervised learning, a program does not learn from labelled data. Instead, it attempts to discover patterns in the data on its own. For example, suppose you have two classes scattered in a 2-dimensional space (as in the first of the images below) and you want to separate the two data sets (as in the second image). Unsupervised learning finds underlying patterns in the data, allowing the classes to be separated.

Unsupervised Learning



Unsupervised Learning

- To highlight the difference between supervised and unsupervised learning, consider the following example.
- Assume that you have collected data describing the heights and weights of people.
- An unsupervised clustering algorithm might produce groups that correspond to men and women, or children and adults.
- An example of a supervised learning problem is if we label some of the data with the person's gender, and then try to induce a rule to predict whether a person is male or female based on their height and weight.

Unsupervised Learning

- **The following algorithms have proven to be highly valuable in practical applications, making them some of the more commonly used methods in unsupervised learning.**

  - **K-Means Clustering**
  - **Hierarchical Clustering**
  - **t-Distributed Stochastic Neighbour Embedding**
  - **Gaussian Mixture Models**
  - **Autoencoders**

CoGrammar

Q & A SECTION

Please use this time to ask any questions relating to the topic, should you have any.

CoGrammar

Thank you for joining us

1. Take regular breaks
2. Stay hydrated
3. Avoid prolonged screen time
4. Practise good posture
5. Get regular exercise

"With great power comes great responsibility"