**Week 7 Pre-Lecture Handout**

**Introduction to Machine Learning: Principles and Overview**

**What is Machine Learning?**

Machine Learning (ML) is a field of artificial intelligence (AI) that enables computers to learn from and make decisions based on data. Unlike traditional software, which follows predefined rules, ML models identify patterns in data and use these patterns to make predictions or automate decision-making processes.

**Why is ML Useful? An Example: Diagnosing Diabetes**

Consider the challenge of diagnosing diabetes. A traditional approach might involve manually defining a set of rules based on domain expertise – for example, through speaking to medical experts. The rules might look something like this (as an example, I’m not a medical expert!):

* If blood sugar > 200 mg/dL and frequent urination is reported, then classify as diabetic.
* If blood sugar is between 140 and 199 mg/dL, consider prediabetes.

However, real-world diagnosis is far more complex. There are numerous factors at play, such as patient history, lifestyle, and genetic predispositions. Defining rigid rules that account for all these factors is impractical. In some cases, we won’t even know how to define such a clear set of rules. ML allows us to leverage large datasets of past diagnoses to train models that can learn intricate patterns and make more accurate predictions.

In **supervised** **machine learning** we use an **algorithm** to learn directly from **training data** to build a **model** which allows us to predict a **target label** based on **features** in the data. This is done by **training** the model, which involves adjusting its **parameters** so that we minimise the error in its predictions. We can then use the model for **inference** in cases where we don’t have the target label.

There’s a lot of terminology used here so let’s unpack this and provide a few key definitions.

**Key Definitions**

* **Algorithm:** A sequence of well-defined instructions or steps that a computer follows to solve a problem or perform a computation. In ML, algorithms are used to train models by learning patterns from data.
* **Model**: A mathematical representation that we can use to make predictions or classifications based on input data.
* **Training Data**: The dataset used to teach a machine learning model by adjusting its parameters.
* **Features**: The independent variables or input variables used by a model to make predictions (e.g., blood sugar levels, age, weight).
* **Labels/Targets**: The dependent variable or the output the model is trying to predict (e.g., whether a patient has diabetes: Yes/No).
* **Training**: The process of adjusting the model's parameters based on data to minimise errors in predictions.
* **Inference**: Using a trained model to make predictions on new, unseen data.
* **Parameters**: Values that the model learns from the data (e.g., the splits in a decision tree or the weights in a neural network).
* **Hyperparameters**: Settings configured before training that affect how the algorithm learns (e.g., tree depth in a decision tree, learning rate in gradient descent).

**Types of Machine Learning**

ML can be broadly categorised into three main types:

**1. Supervised Learning**

In supervised learning, the model is trained on **labelled data**, meaning that each input is paired with a corresponding correct output. For example, we might have a dataset of past patients where each row contains medical features (e.g., blood pressure, insulin levels) along with the **label** (whether the patient was diagnosed with diabetes or not).

**Formulating Real-World Problems as Classification or Regression**

* **Classification**: Assigning an instance to one of several categories. Example: Given a patient's medical history, predict whether they have diabetes (Yes/No). Other examples include email spam detection (spam vs. not spam), image recognition (e.g., cat vs. dog).
  + If the label only has two possible values (e.g., Yes/No, True/False) then this is referred to as binary classification
  + If the label has more than two possible values, then this is referred to as multiclass classification
* **Regression**: Predicting a continuous outcome. Example: e.g., stock price prediction, house price estimation.

We start with a **labelled dataset** (in our example, where we know whether past patients had diabetes or not), train the model on this dataset, and then deploy it in the real world to make predictions on **new patients whose diabetes status is unknown**. This ability to generalise to unseen cases is what makes ML powerful in real-world settings.

**Example: Decision Trees for Diabetes Diagnosis**

A **Decision Tree** is a simple yet powerful supervised learning algorithm. It works by splitting the data into branches based on decision rules (again this is intended as an illustrative example and is not medically accurate!):

1. If blood sugar > 200 mg/dL → Classify as diabetic.
2. Else, if blood sugar between 140 and 199 mg/dL AND high BMI → Classify as diabetic.
3. Else → Classify as non-diabetic.

This is referred to as a decision tree because we can represent these rules as branches on a tree-like structure as illustrated on the next page. To make a prediction, we work from the top of the tree to one of the **leaf nodes**. Once at a leaf node, we make a prediction. So, for example, if we have a patient with blood sugar of 160mg/dL and low BMI we would take the ‘Yes’ branch from the **root node** (the node at the top of the tree) and then the ‘Yes’ branch at the next **branch node** (blood sugar > 140mg/dL) and the ‘No’ branch at the next node (high BMI) to give a prediction of non-diabetic at the **leaf node**. We then have reached a leaf node and here we make the prediction that they are **not** diabetic. If you look at the rules listed above, you can see this is just a visual way of representing these same rules.

A decision tree is therefore a **model** we can use for classification (in this example, for making a diagnosis of a patient given some medical data). The aim of the decision tree machine learning algorithm is to construct the decision tree that best fits the training data.

A diagram of a patient's level

AI-generated content may be incorrect.

**Parameters vs. Hyperparameters**:

* **Parameters**: Parameters are the internal values that a machine learning model learns from the training data. They directly influence how the model makes predictions. Parameters are adjusted automatically during training to minimise error. Different models have different types of parameters. For example, for decision trees, the parameters include the decision rules learned by the model, such as which features to split on and the thresholds to use (e.g., the fact that blood sugar > 200 mg/dL is the optimal threshold for classifying as diabetic
* **Hyperparameters**: Hyperparameters are external settings that control how the model is trained. Unlike parameters, they are not learned from the data but are set manually before training begins. They are chosen by the user or optimised using techniques like grid search or random search. In the context of decision trees, it includes predefined choices like the maximum depth of the tree (i.e., how many decisions are made before classification).

**2. Unsupervised Learning**

In unsupervised learning, the model is trained on **unlabelled data** and must find patterns and structures within the data without predefined outputs.

**Examples of Unsupervised Learning:**

* **Clustering**: Grouping similar items together, e.g., customer segmentation in marketing.
* **Anomaly Detection**: Identifying rare or unusual patterns, e.g., fraud detection in financial transactions.
* **Dimensionality Reduction**: Reducing the number of input variables while retaining essential information, e.g., Principal Component Analysis (PCA).

**3. Reinforcement Learning**

Reinforcement learning (RL) is a type of machine learning where an agent learns to make decisions by interacting with an environment to maximise a cumulative reward.

**Examples of Reinforcement Learning:**

* **Game Playing**: AI learning to play chess or Go.
* **Robotics**: Teaching robots to walk or perform tasks.
* **Autonomous Vehicles**: Self-driving cars optimising their navigation strategies.

**Machine Learning Workflow**

The typical workflow of an ML project involves several key steps:

1. **Problem Definition**: Understanding the business or research question and defining the ML task (classification, regression, etc.).
2. **Data Collection**: Gathering relevant data from various sources.
3. **Data Preprocessing**: Cleaning, transforming, and preparing data for training.
4. **Feature Engineering**: Selecting and constructing relevant features for the model.
5. **Model Selection**: Choosing an appropriate ML algorithm.
6. **Training**: Feeding the training data into the model and optimising its parameters.
7. **Evaluation**: Assessing model performance using validation data and metrics.
8. **Tuning**: Optimising hyperparameters to improve performance.
9. **Deployment**: Integrating the trained model into a production system.
10. **Monitoring and Maintenance**: Continuously tracking performance and updating the model as needed.

**Common Machine Learning Algorithms**

Different ML algorithms serve different purposes. Below are some widely used algorithms for each type of learning:

**Supervised Learning Algorithms:**

* **Linear Regression**: Predicts a continuous value by fitting a straight line to data points.
* **Logistic Regression**: Adaption of regression for classification problems.
* **Decision Trees**: A flowchart-like structure for decision-making.
* **Random Forest**: An ensemble of multiple decision trees to improve accuracy.
* **Support Vector Machines (SVM)**: Finds the optimal boundary between different classes.
* **Neural Networks**: Inspired by biological neurons, these are used in deep learning for complex pattern recognition.

**Unsupervised Learning Algorithms:**

* **K-Means Clustering**: Groups data into K clusters based on similarity.
* **Hierarchical Clustering**: Creates a hierarchy of clusters.
* **Principal Component Analysis (PCA)**: Reduces the number of dimensions in the dataset while preserving important information.

**Applications of Machine Learning**

Machine learning is widely used across industries. Here are some key applications:

* **Healthcare**: Disease diagnosis, drug discovery, and personalised treatment recommendations.
* **Finance**: Fraud detection, credit risk assessment, and algorithmic trading.
* **Marketing**: Customer segmentation, recommendation systems, and sentiment analysis.
* **Retail**: Demand forecasting, inventory optimisation, and chatbots.
* **Transportation**: Route optimisation, self-driving cars, and predictive maintenance.
* **Manufacturing**: Quality control, predictive maintenance, and supply chain optimisation.

**Challenges and Limitations of Machine Learning**

Despite its advantages, ML has several challenges:

* **Data Quality**: ML models rely heavily on the quality and quantity of data.
* **Bias and Fairness**: Models can inherit biases from training data, leading to unfair decisions.
* **Overfitting**: When a model performs well on training data but poorly on new data.
* **Interpretability**: Complex models, especially deep learning models, can be difficult to understand and explain.
* **Computational Requirements**: Training large models requires significant computational resources.

**Ethical Considerations in Machine Learning**

Ethical issues in ML include:

* **Privacy**: Ensuring data security and user privacy.
* **Bias**: Addressing potential biases in model predictions.
* **Transparency**: Making ML decisions interpretable and explainable.
* **Accountability**: Defining responsibility for ML-driven decisions.