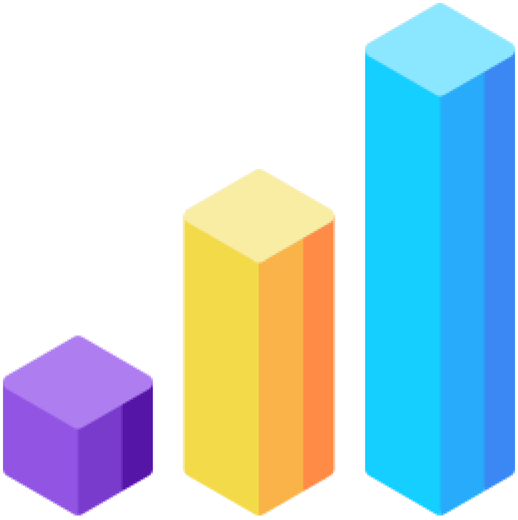

SIT720 Machine Learning

Task 2

Michael Rideout
Student Id: 225065259

Item 1.1 Main Points Summary

Week 1 Main Points Summary

The main points from week 1's lecture can be summarised as follows:

Foundation of ML (ML)

- **Steps in Machine Learning**:
    - Data Collection
    - Clean and Prepare Data
    - Train or Build Model
    - Evaluate the Model
    - Improve/Deploy
- **Varieties of Machine Learning**:
    - **Unsupervised Learning:** Finds the underlying structure in unlabelled data.. Some types include:
        - **Clustering:** The grouping of data that is identified as similar based on some metric.
        - **Anomaly Detection:** Identifying unusual data points.
        - **Data Understanding and Visualisation:** Making sense of and representing data.
        - **Information Retrieval**
        - **Data Compression (Reduction):** Reducing the dimensionality of data.
    - **Supervised Learning:** Aims to discover or learn a function which models the input data to the output target.
        - **Target function:** $f:X \rightarrow Y$.
        - **Examples:** Training data in the form of (x,y) pairs.
        - **Hypothesis:** $g:X \rightarrow Y$ such that $g(x)=f(x)$.
            - **x:** Set of attribute values (feature vector).
            - **y:** Discrete label (classification) or real-valued number (regression).
        - **Classification:** Is the machine learning task of predicting a discrete target or label from the input feature vector.
        - **Regression:** Is the machine learning task which predicts a continuous real-valued number based on input features.It learns the relationship between dependent and independent variables.
    - **Reinforcement Learning:** A method in machine learning whereby an agent learns by interacting with an environment and maximising some form of reward based on its behaviour.
- **Real-world Applications of ML:** ML is used in various domains, including robotics, board games (e.g., AlphaGo), speech recognition (e.g., Siri), computer vision, healthcare, financial market prediction and many more
- **Model Assessment and Selection:**
    - **Model Evaluation:** Determining if a model will accurately predict labels on new data. Usually achieved by the generation of a training and testing dataset often by splitting all available data. Training generates a learned model based

on the training dataset and evaluation is performed on the test dataset using a measurement like accuracy. This process can be repeated with different random splits, and the results are averaged.
    - **Model Selection:** Finding the best model from many possibilities. This involves considering parameters and hyper-parameters. Approaches include looking at averaged evaluation scores and using cross-validation (training on one set and testing on another, then rotating them).
    - **Overfitting:** Is when a model learns the input data too well in that it performs badly when presented with data not in the training dataset. The model isnt sufficient general in this case.

Revising Knowledge of Linear Algebra

- **Vectors:** A data instance is often represented as a feature vector, which is an ordered list of numbers. Vectors can be represented as column or row vectors.
- **Vector Operations:**
    - **Transpose:** This operation changes a column vector into a row vector, or it can change a row vector into a column vector.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \Rightarrow X^T = [x_1 \ x_2 \ \dots \ x_n]$$

    - **Addition:** It performs element-wise addition between two vectors, provided they have an equal number of elements.

$$[X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}][Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}][X + Y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}]$$

    - **Inner Product (Dot Product):** The sum of the products of the corresponding elements of two vectors.

$$XY = [x_1y_1 + x_2y_2 + \cdots + x_ny_n]$$

    - **Magnitude/Length (Norm):** The length of a vector, commonly calculated using the **2-norm** (Euclidean norm). It is $\text{length}(X) = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ also written as $\|X\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$
      The **p-norm** is a generalisation. $\|X\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}$
    - **Similarity:** How alike two vectors are based on distance or similarity measures
        - Euclidean distance is $\|x - y\| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$
        - Cosine similarity is $\cos(\theta) = \dfrac{x^T y}{\|x\|\|y\|} = \dfrac{x^T y}{\sqrt{x^T x}\sqrt{y^T y}}$

- **Matrices:** A matrix consists of a set of numerical elements systematically ordered within a rectangular layout of rows and columns.
    - Mathematically this is $A \in \mathbb{R}^{m \times n}$ where m is the number of rows and n is the number of columns
- **Matrix Types:**
    - **Rectangular Matrix:** Number of rows is not equal to the number of columns (m ≠ n). E.g. $\begin{bmatrix} 1 & 2 & 5 \\ 6 & 2 & 4 \end{bmatrix}$
    - **Square Matrix:** There are the same number of rows as there are columns.(m = n). E.g. $\begin{bmatrix} 1 & 6 \\ 2 & 3 \end{bmatrix}$
    - **Symmetric Matrix:** A square matrix equal to its transpose (A = A^T). E.g. $\begin{bmatrix} 1 & 6 \\ 2 & 3 \end{bmatrix}$
    - **Diagonal Matrix:** A square matrix where all off-diagonal elements are zero (A(i,j) = 0 for i ≠ j). E.g. $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 5 \end{bmatrix}$
    - **Identity Matrix:** It's a diagonal matrix where every value along the main diagonal is exactly 1. (I(i,i) = 1). E.g. $I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
- **Matrix Operations:**
    - **Transpose:** Obtained by interchanging rows and columns. E.g.
      $$A = \begin{bmatrix} 1 & 6 & 7 \\ 2 & 3 & 8 \end{bmatrix} \implies A^T = \begin{bmatrix} 1 & 2 \\ 6 & 3 \\ 7 & 8 \end{bmatrix}$$
    - **Addition/Subtraction:** Applying addition or subtraction operations individually to corresponding entries. Requires the matrices to be of equal size. E.g. $X + Y = \begin{bmatrix} 2 & 4 \\ 3 & 1 \\ 8 & 5 \end{bmatrix} + \begin{bmatrix} 6 & 7 \\ 4 & 4 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 8 & 11 \\ 7 & 5 \\ 9 & 8 \end{bmatrix}$
    - **Scalar Multiplication/Division:** It involves taking a single numerical value (the scalar) and applying multiplication (or division) with this value to every individual entry within the matrix.r. E.g. $3 \begin{bmatrix} 6 & 7 \\ 4 & 4 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 18 & 21 \\ 12 & 12 \\ 3 & 9 \end{bmatrix}$
    - **Element-wise Matrix Multiplication:** It involves taking two matrices of identical dimensions and creating a new matrix where each entry is the result of multiplying the entries found at the same position (same row and column) in the original two matrices.e. E.g. $\begin{bmatrix} 2 & 4 \\ 3 & 1 \\ 8 & 5 \end{bmatrix} \odot \begin{bmatrix} 6 & 7 \\ 4 & 4 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 12 & 28 \\ 12 & 4 \\ 8 & 15 \end{bmatrix}$

- **Matrix to Matrix Multiplication:** Standard matrix multiplication requires compatible dimensions: the number of columns in the leading matrix must equal the number of rows in the trailing matrix. The resulting element C(i,j) is the dot product derived from the i-th row of the first matrix and the j-th column of the second. Notably, this multiplication is non-commutative, meaning AB ≠ BA in most cases.. E.g. $\begin{bmatrix} 2 & 4 \\ 5 & 6 \\ 1 & 7 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 16 \\ 11 & 28 \\ 8 & 23 \end{bmatrix}$
- **Inverse Matrix:** Two square matrices, A and B, are inverses of each other provided their product, irrespective of the multiplication order (AB or BA), is the identity matrix I. We denote A as the inverse of B using B^{-1}. A square matrix possesses an inverse if, and only if, it has a non-zero determinant. Matrices with a zero determinant do not have an inverse.

Week 2 Main Points Summary

The main points from week 2's lecture can be summarised as follows:

Linear Algebra: Feature Vectors and Matrices

- **Feature vector**: mathematically depicts a group of documents as a vector, is an essential, basic procedure within information retrieval.
- **Vocabulary of features**: A standard practice for text data involves generating a list of relevant features (a vocabulary) that covers all instances present in the dataset.
- **Vocabulary:** The vocabulary serves as a template to represent every instance as a distinct feature vector. Therefore, the number of vectors produced will equal the number of instances, N.
- A **feature matrix** X is formed by stacking these feature vectors as columns (or rows).

Probability Concepts

- **Random Experiment:** process yielding an outcome that lacks certainty prior to its occurrence. Consider, for instance, the act of tossing a coin or rolling dice.
- Whereas the sample space (Ω) lists the totality of conceivable results for a random experiment, an event constitutes just a particular set (or grouping) of those results.
- **Probability** $P(A)$ is the numerical quantification of the likelihood that a given event A will occur, with this value always falling in the range [0,1] The probability of an event A not occurring is $P(\overline{A}) = 1 - P(A)$.
- **Joint probability** is the probability of more than one event occurring. If two events A and B are independent, their joint probability is $P(A \text{ and } B) = P(A) * P(B)$

- **Conditional probability** $P(A|B)$ is the probability of event A occurring given that event B has occurred. It is defined as: $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$, provided $P(B) \neq 0$.
- **Bayes' Rule** describes how to update existing beliefs in light of new evidence. It states: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$, where $P(B) \neq 0$

Random Variable

- A **random variable**'s value is determined by the outcome of a random event or experiment. Think of it as a function that assigns a probability score to each relevant event associated with that experiment.
- **Discrete Random Variables** have a countable number of values (e.g., faces of a dice).. They are defined using a **Probability Mass Function (PMF)** $\pi(x) = P(X = x)$, where $\sum_x \pi(x) = 1$. The **Cumulative Distribution Function (CDF)** $F(X)$ gives the cumulative probability $F(X) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$.
- **Continuous Random Variables** can take values on an infinite continuum (e.g., height). They are defined using a **Probability Density Function (PDF)** $f(x)$ such that the probability over a range $[a, b]$ is $P(a \leq X \leq b) = \int_a^b f(x)dx$, and $\int_{-\infty}^{+\infty} f(x)dx = 1$. The probability at any exact value is zero.

Distribution of Random Variables

- A **probability distribution** serves as a mapping, connecting each potential outcome of a statistical experiment to its particular chance of happening.
- **Bernoulli Distribution** is a discrete distribution for a binary random variable ($X = 0$ or $X = 1$) with $P(X = 0) = p$ and $P(X = 1) = 1 - p$. Notation: $x \sim B(x|p)$.
- **Uniform Distribution**:
  - Discrete: All values are equally likely, $P(X = x_i) = \frac{1}{N}$ for $i = 1..N$. Notation: $x \sim U(x|N)$.
  - Continuous: $f(x) = \frac{1}{b - a}$ for $a \leq x \leq b$]. Notation: $x \sim U(x|a, b)$.
- **Normal Distribution** is a continuous distribution defined by: $N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2\sigma^2}(x - \mu)^2$]. Notation: $x \sim N(x|\mu, \sigma^2)$. It is popular as many natural phenomena approximately follow this distribution.
- The **Central Limit Theorem** holds that when you calculate the means from multiple sufficiently large samples drawn from any population, the distribution of those sample

means will closely resemble a normal distribution, regardless of how the original population's data is distributed. The same principle applies to the sum of many independent and identically distributed random variables.

Data Wrangling

- **Data wrangling** (or munging) is the process of cleaning, transforming, and organising a dataset to make it suitable for analysis.
- **Missing value replacement** is essential as machine learning models cannot process null values. Methods include replacing with the immediate value, mean, or median of the row/column.
- **Scaling or normalisation** transforms the values of a dataset into a common range to improve the performance of machine learning algorithms
- **Min-max normalisation** scales data to a range of 0 to 1 using the formula:
  $$v' = \frac{v - \min(v)}{\max(v) - \min(v)}.$$
- **Non-numeric data encoding** converts categorical values into integer values that machine learning models can understand. Techniques include **OrdinalEncoder**, **One-Hot Encoding**, and **LabelEncoder**.

1.2 Summary of Reading List Items

Week 1 Readings

1.4 Defining Machine Learning

Wikipedia 'Machine Learning' Article - https://en.wikipedia.org/wiki/Machine_learning

Machine learning (ML) is a field within artificial intelligence where statistical algorithms learn from data to perform tasks without explicit instructions, generalizing to new, unseen data. It encompasses various approaches, including:

- Supervised learning: Algorithms learn from labeled data (inputs and desired outputs) to create a model that maps inputs to outputs. This includes classification (predicting categories) and regression (predicting numerical values).
- Unsupervised learning: Algorithms find structure and patterns in unlabeled data. Key applications are clustering (grouping similar data) and dimensionality reduction.
- Reinforcement learning: Algorithms learn by interacting with an environment, receiving feedback (rewards or penalties) to achieve a specific goal.

1.14 Matrix Algebra

Introduction to Eigenvalues
https://www.youtube.com/watch?v=G4N8vJpf7hM&ab_channel=PatrickJ

This video introduces eigenvalues and eigenvectors. It defines them, provides a simple example demonstrating the relationship between a matrix (A), an eigenvector (x), and an eigenvalue (λ) where $Ax = \lambda x$, and mentions that multiple eigenvectors can be associated with a single eigenvalue.

Finding Eigenvalues and Eigenvectors
https://www.youtube.com/watch?v=IdsV0RaC9jM&ab_channel=PatrickJ

This video tutorial demonstrates how to find the eigenvalues and eigenvectors of a 2x2 matrix. The process involves calculating the determinant of ($A - \lambda I$), where A is the matrix, λ represents eigenvalues, and I is the identity matrix. Solving for λ gives the eigenvalues, which are then substituted back into the ($A - \lambda I$) matrix to solve for the corresponding eigenvectors through row reduction.

Eigenvectors and eigenvalues
https://www.youtube.com/watch?v=PFDu9oVAE-g&ab_channel=3Blue1Brown

This video explains eigenvectors and eigenvalues, clarifying their meaning and application. It emphasizes the importance of a strong foundational understanding of linear transformations, matrices, determinants, and change of basis before tackling this topic. The core concept is that eigenvectors are special vectors which, when transformed by a matrix, only scale (stretch or shrink) and don't rotate. The scaling factor is the eigenvalue. It details how to compute eigenvectors and eigenvalues, linking the process to the determinant of a modified matrix and showing how an eigenbasis (a basis formed by eigenvectors) simplifies calculations, particularly when raising a matrix to a high power.

Week 2 Readings

Understanding Random Variables
https://www.youtube.com/watch?v=lHCpYeFvTs0&ab_channel=DrNic%27sMathsandStats

This video explains random variables, differentiating between discrete and continuous variables using the example of an ice cream stand's sales data. It demonstrates how to calculate probabilities using a discrete distribution and highlights the difference between data suitable for discrete and continuous modelling..

Python Libraries

Python libraries utilised in this task include:

- **Random:** Standard library used for generating pseudo-random numbers
- **Hashlib:** Suite of hashing algorithms
- **Tabulate:** Plain text tables
- **Pandas:** Data manipulation and analysis library
- **Numpy:** Library for scientific and numerical computing
- **Mathplotlib:** Visualisation library
- **Seaborn:** Visualisation library
- **Sklearn:** Machine learning toolkit

1.3 Learning Reflection

Based on the learning outcomes of SIT720 Machine Learning as described in the university handbook (https://www.deakin.edu.au/courses-search/unit.php?unit=SIT720) the following objects were achieved in weeks 1 and 2:

ULO1: Maintain in-depth knowledge of advances in machine learning, and use this knowledge to explain machine learning techniques and algorithms to a range of technical and non-technical audiences.

Weeks 1 and 2 of SIT720 Machine Learning provided a solid foundation in core machine learning concepts, enabling the development of in-depth knowledge (GLO1). Key areas covered include:

- **Fundamentals:** Defining the typical steps in a machine learning process, from data collection and preparation to model training, evaluation, and deployment
- **Learning Paradigms:** Differentiating between major types of machine learning:
    - **Supervised Learning:** Learning a mapping from inputs (X) to outputs (Y) using labeled data ((x,y) pairs) encompassing tasks like classification (predicting discrete labels) and regression (predicting continuous values)
    - **Unsupervised Learning:** Discovering patterns in unlabeled data, including techniques like clustering, anomaly detection, and dimensionality reduction.
    - **Reinforcement Learning:** Agents learning through interaction with an environment to maximize rewards.
- **Core Concepts:** Understanding essential ideas like feature vectors, model evaluation, overfitting (where a model learns training data too well but fails to generalize)], and the importance of generalization.
- **Mathematical Foundations:** Revisiting crucial linear algebra concepts (vectors, matrices, operations like transpose, dot product, inverse, eigenvalues/eigenvectors) and probability theory (random variables, distributions like Normal, Bernoulli, Uniform, Bayes' Rule, Central Limit Theorem, which underpin many ML algorithms.

ULO2: Explore data using a range of machine learning techniques, evaluate resulting models, and extract and communicate insights from data in real-world scenarios.

- As part of Task 2 weather data was analysed and manipulated using numpy and pandas.

ULO3: Justify proposed solutions by evaluating and comparing results from alternative approaches to solving real-world problems and exploring data using machine learning techniques.

- **Model Selection:** Introducing the concept of selecting the best model from many possibilities, considering parameters and hyperparameters.
- **Evaluation Strategies:** Highlighting evaluation using train/test splits and cross-validation as methods to compare model performance.
- **Comparing Approaches:** Understanding the different learning paradigms (supervised, unsupervised, reinforcement) allows for reasoning about which approach might be most suitable for a given real-world problem (e.g., using classification for spam detection vs. clustering for customer segmentation).
- **Critical Evaluation:** Awareness of potential pitfalls like overfitting encourages critical thinking about model performance and the need for robust evaluation before proposing a solution.

ULO4: Create Python scripts to automate the evaluation and analysis of data using a range of machine learning libraries, techniques, and algorithms.

- Task 2 required basic data manipulation and analysis using Pandas and numpy

1.4 Quiz Results

Week 1 quiz:

Week-1 quiz

Your work has been saved and submitted

Written 29 March, 2025 10:21 PM - 29 March, 2025 10:23 PM • Attempt 2 of unlimited

Your quiz has been submitted successfully, the answer(s) for the following question(s) are incorrect.

Attempt Score **9 / 10 - 90 %**

Overall Grade (Highest Attempt) **9 / 10 - 90 %**

Week 2 Quiz:

Week 2 quiz

×

Your work has been saved and submitted

Written 30 March, 2025 1:59 AM - 30 March, 2025 2:06 AM • Attempt 1 of unlimited

Your quiz has been submitted successfully, the answer(s) for the following question(s) are incorrect.

Attempt Score	**9 / 10 - 90 %**
Overall Grade (Highest Attempt)	**9 / 10 - 90 %**