6G6Z0048 Artificial Intelligence: 1CWK100

Delete all the reminder text in grey before you submit.

.

# 1) Longlisting [~1,500 words]

**1792**

**Supervised Learning** is the process of training a predictive model as is it given clear instructions on what needs to be learnt and how to learn it.

**unsupervised learning** is the process of training a descriptive model as there is no single feature or particular interest and no target to learn.

**K-Nearest Neighbour** (kNN) is a simple learning algorithm that classifies unlabelled examples using its k-nearest neighbour’s information. Using the features of the examples as coordinates, it plots the examples in a feature space. To classify unlabelled examples, the distance between the example and other labelled examples is calculated using the Euclidean distance method. The unlabelled example is then classified in the same class as the closest labelled example. K is the hyperparameter of the number of neighbours that are compared to the example, this can be altered to improve the accuracy of model. kNN can be applied to image classification. For example, the user uploads an image of a bike, the algorithm will then highlight features of the image, i.e. “wheels” and use the features to compare it to existing labelled images in the feature space. It will then label the new image with its nearest neighbour which is bike.

**k-means clustering** assigns one of the examples to one of the k clusters, where k is a number that has been determined ahead of time. The goal is to minimise the difference in feature values of examples within each cluster and maximise the difference between each cluster. As it is Unsupervised Learning, it starts with an initial guess for cluster assignments and modifies slightly to see for changes that improve homogeneity within clusters. It stops once there is no change is reassignment of examples. In image classification, the images can be assigned to clusters and modified until no changes are made, to train the model. New images can then be classified by the trained model.

**Feature Space Visualisation** is used to represent images as points in a multi-dimensional feature space. Within the feature space lies all possible combination of feature values (Lantz, 2023). Feature space visualisation helps to find patterns and relationships between data, allowing them to be grouped in the similar classes/clusters.

**Hyperparameter Tuning** is the testing of hyperparameter settings to achieve the better model fit. For example, tuning k in kNN to find the best value of k, which is the number of neighbours compared to the unlabelled example.

**Naïve Baye Classification** uses training data to calculate the probability of each outcome based on evidence provided by feature values. It uses calculated probability to predict the class of a new example. Naïve baye assumes that all features in a dataset are equally important and independent (Lantz, 2023). Due to the algorithm using frequency tables to learn the data, each feature must be categorical therefore numerical features must be put into categories known as bins. Naïve Baye requires features to be extracted before the probability can be determined. One approach to this is to use Bag of Visual words, which treats features as words and displays them in a histogram based on how many times it appears. This can be applied to image classification as BoVW would use the image features as words to create a histogram. For example, BoVW would find the words “wheel”, “frame”, “handle”, which are then used by Naïve baye. If a new image had these words, Naïve baye would use the probability to classify this image as bike.

**Bag Of Visual Words** represents data as a binary feature indicating whether each word appears in an example. It allows images to be directly comparable points in the same feature space.

**Decision Tree Classification** uses a tree structure to model the relationship between features and the possible outcomes. Examples start at the root node and pass through the decision tree based on its features. Each node branches off into more than one node. The example continues through the decision tree until it ends at a leaf/terminal node, which will be the class for the example. The decision tree is a divide and conquer algorithm as it divides the dataset into subsets which are split repeatedly until the process stops as the data in the subsets become homogenous. In Image classification, the example of a bike image will start at the root node of the tree, the first node could be “Does the image have wheels?”. The image will move to the YES node. This will continue until the image has reached the terminal leaf which will classify the image as bike.

**Gradient boosting**

**Ensemble Classification** combines multiple weak learners to create a stronger learner. The ensemble depends on a diverse set of classifiers as their individual predictions are combined into one final prediction. When each classifier makes its own independent prediction, it must do more than simply guess. This means that ensemble has an uncorrelated classification but is better than random chance. Ensemble classification could combine multiple weak image classification learners to create a stronger learner. Each learner would independently decide which class the image should belong to, and the final decision is determined by majority vote.

**Boosting** is an ensemble method, which improves the performance of weak learners to attain the performance of strong learners. Boosting uses ensembles of models trained on resampled data, which vote to determine the final prediction. The prediction is biased as models that perform better have a greater influence over the ensemble’s final prediction. This can improve overall accuracy of image classification as it allows the better performance models to have greater influence over the final decision.

**Bagging** Is an ensemble method which generates new training data using bootstrap sampling on original training data (Lantz, 2023). Datasets are used to generate a set of models using single learning algorithm. Similar to boosting, the models’ predictions are combined via voting for classification. Bagging is often used with unstable learners; these are essential for ensuring diversity as they vary dramatically if minor changes occur in the input data. For image classification, the models are trained on a different subset of images.

**Random Forests** builds upon bagging but adds diversity by allowing algorithms to choose from a randomly selected subset. For example, at the root node it might only be allowed to choose from a small number of features, which are chosen at random from a set of predictors at each split. There is always a random different subset provided so each tree in the ensemble is unique. Similar to bagging, once the ensemble of trees is made, it performs simple votes to make a final prediction. For image classification, each tree will be specialised in different aspects or patterns of data. The relevant features of the image are extracted, and each tree in the forest focuses on a different aspect of the image. Like decision trees, the data will start at the root node and make its way through the tree until a terminal node is reached. Once each tree has made its independent prediction, the overall prediction is the class that was most predicted.

**Logistic Regression** is used to calculate the probability of a binary event occurring (Thanda, 2023). It predicts the probability based on a set of independent variables, which influence the outcome. Logistic regression uses the sigmoid function to model the relationship between the input and probability of outcome. It has a decision boundary that splits the instances of the two classes within the feature space. Examples are classified by which side of the decision boundary they lie upon. In image classification, the features of the image are used as independent variables to find the probability of the class, however, as there are only two classes the image is either true or false. For example, the two classes could be “bike” or “not bike”.

**Multinomial Logistic Regression** is an extension of logistic regression and is used when the outcome being predicted has more than two categories (Libguides: Statistics Resources: Multinomial logistic regression). It uses the SoftMax function to turn a vector of K real values into a vector of K real values that sum to 1 (Wood, 2019). The decision boundary is a hyperplane in the feature space that separates the classes. In image classification, multinomial logistic regression will use the features of the image as independent variables to find where it lies within the feature space, and then calculates the probability of it belonging to the class.

**Artificial Neural Network (ANN)** models the relationship between input and output signals. It is like the human brain’s neural network. ANN uses a network of artificial nodes to solve challenging problems. The activation function is the mechanism by which artificial nodes process incoming information and determine whether to pass the signal to the other nodes in the network. If the activation function threshold is met by the input, then it results in an output signal. Each pixel of an image is treated as a separate feature, and the artificial neural network discovers the relationship between these features to make predictions.

**Convolutional Neural Network (CNN)** is a larger deep learning model with more than one hidden layer. The nodes in one layer are only partially connected to the nodes in the next layer. It is a deep feed-forward network, which is used for visual tasks that independently learn important distinguishing image features. In image classification the convolutional layer extracts the features of an image by scanning through the image with filters (Huellmann, 2022). The Rectified Linear Units (ReLU) layer changes any negative values to zero to introduce non-linearity into the network. The pooling layer downsizes the image to increase computational speed, and the fully connected layer outputs a final classification.

**Feed-forward Network** are networks which feed the input signal continuously in one direction from the input layer to the output layer.

**Transfer Learning** uses existing models that were shared, which are deep learning models that can be adapted from one context to another.

**Fine Tuning** is the additional training that a pretrained neural network undergoes if they do not transfer directly to the new task

**Feature Extraction** is a dimensionality reduction technique that capture the relevant features and highlight patterns.

**Support Vector Machine (SVM)** uses multidimensional surfaces to define relationship between features and outcomes. The goal is to create a flat boundary called hyperplane, which divides space to create homogenous partitions on either side. SVM combines kNN and linear regression. It is adaptable for use with nearly any type of learning task. Support vector provides a compact way to store a classification model. In image classification SVM finds the optimal hyperplane that separates the image classes so, when a new image is added it can be classified by mapping it in the hyperplane, and which sector it is in is the class it belongs to.

# 2) Analysis [~2,000 words]

**- about 650 words for each algorithm**

“[…] *critically analysing the suitability of (combinations of) options set out in (1) for adoption by the company, based on:*

1. *the contextual information surrounding this specific image classification problem (see* ***Appendix A****);[[1]](#footnote-1)*
2. *the theoretical characteristics of the various options set out in (1) and/or your own experimental investigations into the options set out in (1) based on a suitable dataset/(s)*”

List good and bad that matter to the 2 founders you ae thinking about

Talk about the founders

Founder 1 – efficient(lots of users); simple (implement in-house), good open source options

Founder 2 – accuracy not critical (only good top 10)

Founder 3 – minimise computer requirement on backend (push to user devices?)

Founder 4 – consistent user experience regardless of object/ appearance (balanced classes?)

founder 1:

She would like a simple and efficient solution to the problem to ensure: i) it can scale; ii) she and the

company retain a good understanding of the codebase, and therefore the ability to maintain, modify and extend it, as

needed. she has started by working with raw pixel values as predictive features, and coding up a kNN classifier.

founder 2:

good top-10 accuracy

founder 3:

interested in any potential there might be for moving computational load away from the back-end, towards the frontend, and onto users’ own devices.

founder 4:

concerned about managing any potential risks, biases or other limitations that might be associated with deploying an AI-based solution. adopt the ‘FAST Track Principles’ of Fairness, Accountability, Sustainability and

Transparency. commonly sold items can belong to object categories for which they presently hold relatively little data, and/or which differ in their appearance from the majority of their existing images.

1. **137 – 173** decision tree - founder 4 transparent for legal reasons. page 138-139.

strengths:

All-purpose classifier that does well on many types of problems

highly automatic learning process, which can handle numeric or nominal features, as well as missing data

excludes unimportant features

can be used on both small and large datasets

results in a model that can be interpreted without a mathematical background

more efficient than other complex models

weaknesses:

DT models are often biased towards splits on features having a large number of levels

it is easy to overfit or underfit the model

can have trouble modelling some relationships due to reliance on axis-parallel splits

small changes in training data can result in large changes in decision logic

large trees can be difficult to interpret and the decisions they make may seem counterintuitive

1. **250 – 267** svm Support Vector Machine (SVM) can be imagined as a surface that creates a boundary between points of data plotted in multidimensional space representing examples and their feature values. Goal is to create flat boundary called hyperplane, which divides space to create homogeneous partition on either side. SVM learning combines aspects of instance based Nearest neighbour and linear regression modelling. This combination allows SVM to model highly complex relationships.

Linearly separable – separated perfected by a straight line or flat surface. Maximum margin hyperplane (MMH) creates the greatest separation between 2 classes. Maximum margin will improve chance that even if random noise is added, each class will remain on its own side of boundary.

Support vectors are the points from each class that are the closes to the MMH. MMH must have at least 1 support vector, which provide a compact way to store classification model.

Linearly separable data: MMH is far away as possible from outer boundaries of 2 groups of data. Outer boundaries are known as convex hull. MMH is perpendicular bisector of shortest line between the 2 convex hulls.

Nonlinearly separable data: Slack variable creates a soft margin that allows some points to fall on the incorrect side of the margin.

SVM has ability to maps problem into higher dimension space using process known as kernel trick, nonlinear relationship may appear linear. Involves process of constructing new feature that express mathematical relationships between measured characteristics.

Strengths: used for classification or numeric prediction problems, not overly influenced by noisy data and not prone to overfitting.

Weakness: slow to train, finding best model requires testing of various combinations of kernels and model parameter.

Strengths:

can be used for classification or numeric prediction problems

not overly influenced by noisy data and not very prone to overfitting

may be easier to use than neural networks, particularly due to existence of several well- supported SVM algorithms

gained popularity due to their high accuracy and high-profile wins in data mining competitions

weaknesses:

finding the best model reequires the testing of various combinations of kernels and model parameters

can be slow to train, particularly if the input dataset has a large number of features or examples.

results in a complex black-box model that is difficult, if not impossible, to interpret

1. **236, 532 – 535** Cnn convolutional layer places early in newtork, comprise most computationally intensive step in network, only layers to process raw image data directly: passes raw data through filter, creating tiles which represent small, overlapping portions of full area.

pooling layer, gather output signal from cluster of neurons in 1 layer and summarise into single neuron for next layer.

These 2 layers serve purpose of identifying important features of images to b learned, reducing dimensionality of dataset.

fully connected layer - used near end of CNN to build model that makes predictions.

designed for image classification.

Input Layer:

Represents the raw pixel values of the input image.

Convolutional Layers:

Perform feature extraction by applying filters (kernels) to detect patterns like edges and textures. Multiple convolutional layers are often stacked to capture hierarchical features.

Activation Function:

Applied after each convolutional operation to introduce non-linearity. Common activation functions include Rectified Linear Unit (ReLU) or variants.

Pooling Layers:

Down sample the spatial dimensions of the feature maps, reducing computational complexity while preserving important information. Common pooling operations include max pooling.

Flattening Layer:

Converts the 2D feature maps into a 1D vector, preparing the data for fully connected layers.

Fully Connected Layers:

Process the flattened features, capturing complex relationships across the entire image. These layers are similar to those in traditional neural networks.

Output Layer:

Produces the final predictions. For image classification, the number of neurons in the output layer corresponds to the number of classes, and the softmax activation function is often used.

# 3) Recommendation [~500 words]

“[…] *drawing on (2) to present a conclusion that argues for a single overall approach to the problem that you believe the company should pursue, and giving your reasons why. No single solution is perfect and this will involve acknowledging weaknesses as well as highlighting strengths.*”

# Reference list

Huellmann, T. (2022) *Image processing: How do image classifiers work?*, *RSS*. Available at: https://levity.ai/blog/how-do-image-classifiers-work#:~:text=The%20four%20types%20of%20CNN,through%20the%20image%20with%20filters. (Accessed: 17 January 2024).

*Libguides: Statistics Resources: Multinomial logistic regression* (no date) *Multinomial Logistic Regression - Statistics Resources - LibGuides at Northcentral University*. Available at: https://resources.nu.edu/statsresources/Multinomiallogistic (Accessed: 17 January 2024).

Thanda, A. (2023) *What is logistic regression? A beginner’s guide [2023]*, *CareerFoundry*. Available at: https://careerfoundry.com/en/blog/data-analytics/what-is-logistic-regression/ (Accessed: 16 January 2024).

Wood, T. (2019) *Softmax function*, *DeepAI*. Available at: https://deepai.org/machine-learning-glossary-and-terms/softmax-layer (Accessed: 17 January 2024).

Add book references from other document

# Appendix 1: LLM conversations

This section should include a straight text-based copy/paste (not images/screengrabs) of any large language model (LLM) conversations you have had regarding the assignment. Formatting doesn’t matter – this section is expected to be long and relatively poorly formatted; all that matters is that the raw text (not images of the text, or similar) is included.

.

# Appendix 2: Source code

This section should include a straight text-based copy/paste (not images/screengrabs) of the full source code for any experimental results you present in the main body of the report. Formatting doesn’t matter – this section is expected to be long and relatively poorly formatted; all that matters is that the raw text (not images of the text, or similar) is included.

.

1. *Note that trying to write section 2 without first reading, and thinking carefully about, the various company information and stakeholder perspectives in* ***Appendix A*** *is likely to seriously limit your mark. Spend time reading the appendix as your first step.* [↑](#footnote-ref-1)