Michael Christopher – S224830467

In the third week, the concept of unsupervised learning was discussed, with a particular emphasis on clustering methods. We studied how clustering techniques are supported by distance metrics (Euclidean, Manhattan, etc.) and how they define similarity between data points. The week spent much of it discussing K-means clustering, including how it functions, how to assess its effectiveness, and the real-world issues like choosing K and initialization sensitivity. An enhancement over the conventional K-means was proposed: K-means++. In order to provide insight into when alternate algorithms could be more appropriate, the week ended with a comparison of different clustering techniques, such as DBSCAN and Hierarchical Clustering.

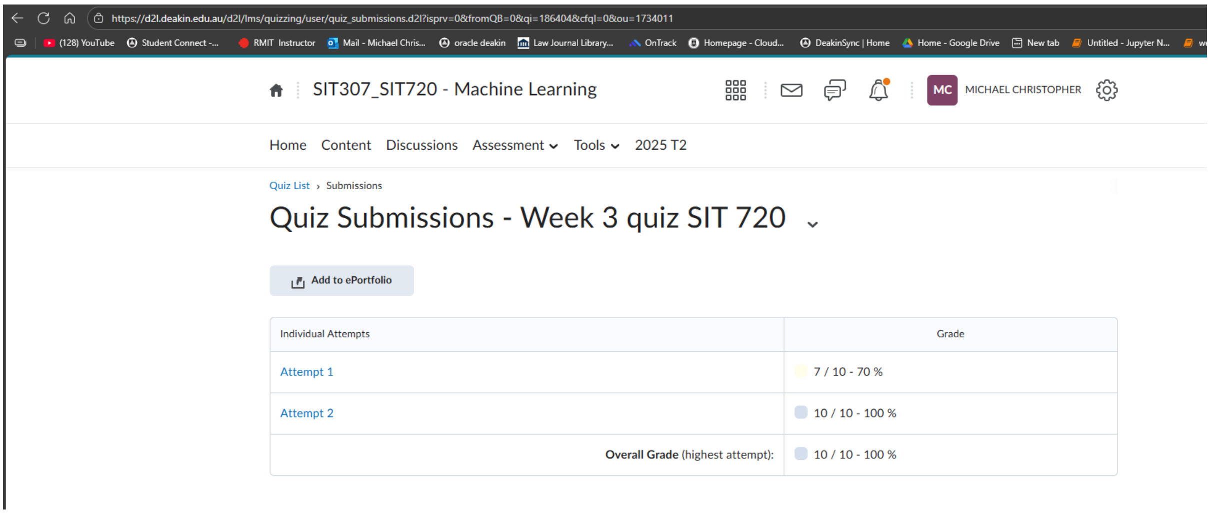
The fourth week was dedicated to dimensionality reduction, beginning with the "curse of dimensionality", the difficulties associated with high-dimensional data. We investigated Principal Component Analysis (PCA), its mathematical foundation, real-world applications, and use cases, especially in facial image analysis, to address these problems. Additionally, the week presented eigenvalue decomposition and singular value decomposition (SVD), which are crucial for comprehending how PCA changes data. Finally, we took a quick look at non-linear methods for lowering dimensions while maintaining data structure, such as t-SNE (t-Distributed Stochastic Neighbour Embedding).

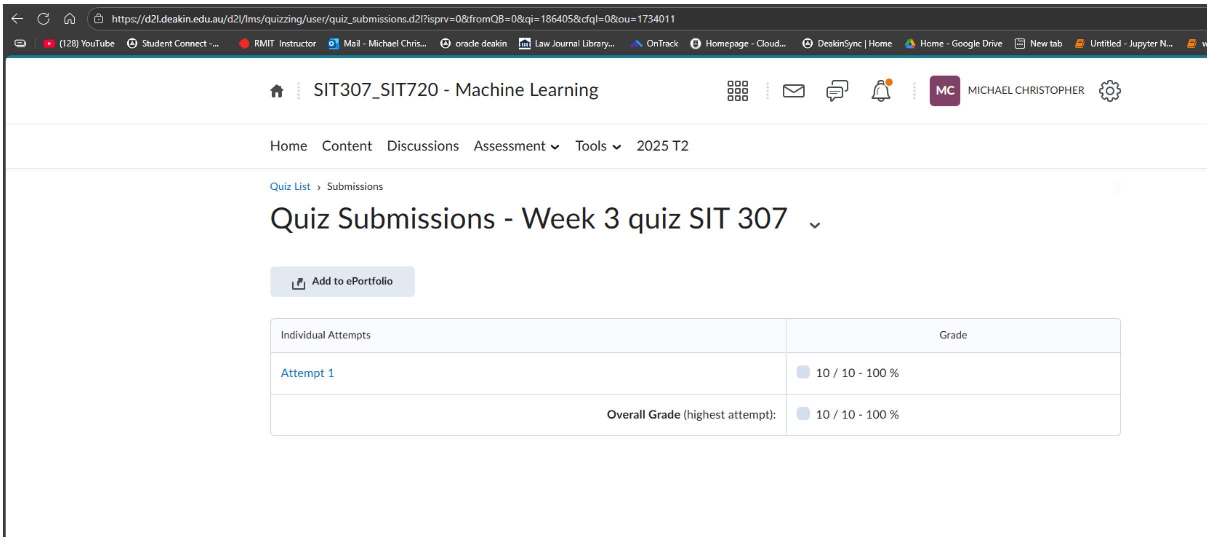
To add extra learning on top of the lecture material, I reviewed the following resources:

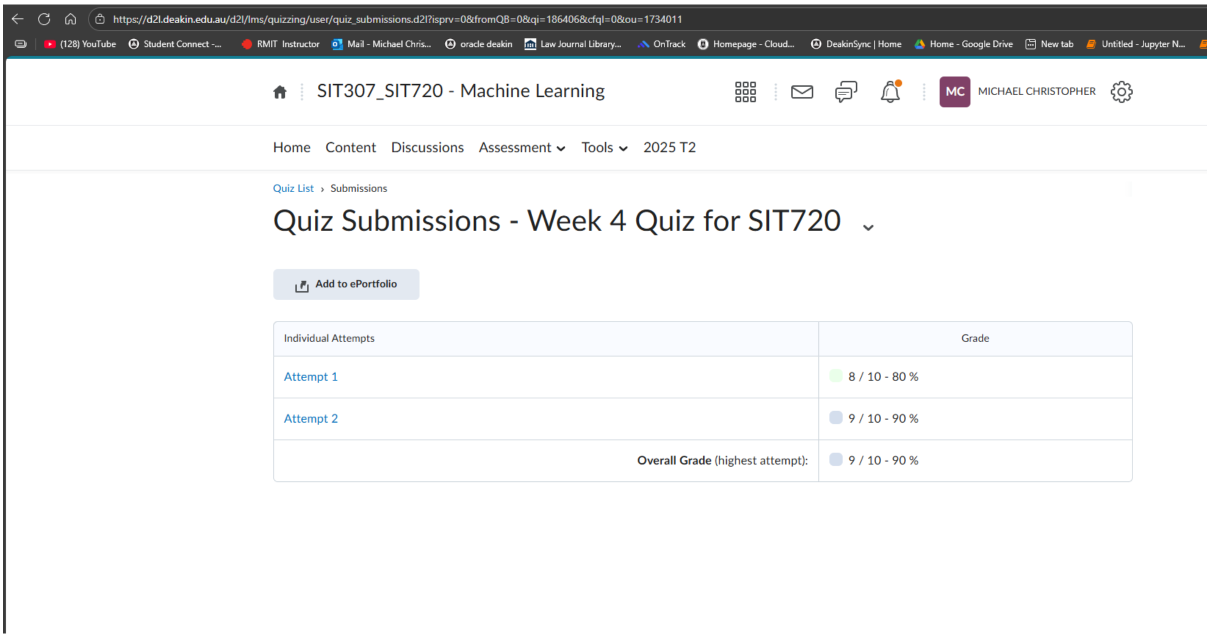
- **Scikit-learn documentation** – Helped me understand practical implementation of K-means, DBSCAN, and PCA in Python.
- **YouTube videos** – Particularly helpful for demystifying eigenvectors, eigenvalues, and PCA using visual and intuitive explanations.

My understanding of unsupervised learning and the mathematical basis for it has greatly increased over the past two weeks. I now see how crucial distance measures are in determining results, whereas previously I believed clustering was solely algorithmic. Additionally, I now have a better understanding of the trade-offs between model complexity and interpretability. Understanding K-means' flaws particularly regarding identifying non-spherical clusters, helped me to see the need for alternative algorithms like DBSCAN.

Although I faced a mathematical obstacle in week four, I gained a great understanding of how linear algebra helps dimensionality reduction. Particularly noteworthy was PCA, both as a method for exposing structure in high-dimensional data and as a tool for data compression. I now see how performance can be enhanced and models can be easier to read by lowering dimensions. All things considered, these weeks have greatly improved my understanding of the mathematical foundations and real-world implications of working with complex, unlabelled datasets in machine learning.







https://d2l.deakin.edu.au/d2l/lms/quizzing/user/quiz_submissions.d2l?isprv=0&fromQB=0&qi=186407&cfql=0&ou=1734011

(128) YouTube | Student Connect -... | RMIT Instructor | Mail - Michael Chris... | oracle deakin | Law Journal Library... | OnTrack | Homepage - Cloud... | DeakinSync | Home | Home - Google Drive | New tab | Untitled - Jupyter N... | we

SIT307_SIT720 - Machine Learning

MC MICHAEL CHRISTOPHER

Home Content Discussions Assessment Tools 2025 T2

Quiz List › Submissions

Quiz Submissions - Week 4 Quiz for SIT307

Add to ePortfolio

Individual Attempts	Grade
Attempt 1	10 / 10 - 100 %
Overall Grade (highest attempt):	10 / 10 - 100 %