CLD 2 Summary Post October 2023: Will Bolton

During this discussion, I gained understanding of the strengths and weaknesses of various ML algorithms, the criteria used to select the most appropriate algorithm for a particular problem, and took the opportunity to understand legal, ethical and technical issues. Deployment of ML algorithms in healthcare poses specific challenges technically and ethically, due to the nature of the data and potential impact on society (Yu et al, 2018; Qayyum et al, 2020). I have included examples of healthcare specific applications throughout my discussion pieces.

Criteria for selecting an appropriate algorithm includes understanding the dataset available (quantity, quality and type of data) and considering the analysis outcomes (Bell, 2020; Russell & Norvig, 2021). Algorithms are broadly classified according to the approaches by which they are designed and operate. These include supervised (e.g. linear or logistic regression, decision trees, support vector machines etc) or un-supervised (e.g. K-means clustering, hierarchical clustering, principle component analysis etc) (Alloghani et al, 2020). A developer may choose supervised learning when using labelled data to make regression predictions or classifications (Alloghani et al, 2020; Bell, 2020; Russell & Norvig, 2021). Developers often choose unsupervised learning approaches due to necessity (lack of labelled data due to time/cost or inherent inability to label), or when they wish to solve problems involving data clustering (Alloghani et al, 2020; Bell, 2020; Russell & Norvig, 2021).

Broad strengths of supervised ML algorithms include predictive accuracy, interpretability (especially with decision trees), and clearer performance metrics (Singh et al, 2016). Weaknesses include the dependency on high quality and quantity of labelled data, overfitting (e.g., capturing noise instead of the underlying pattern), and bias (underfitting) or variance issues. For unsupervised ML algorithms, strengths include their ability to discover hidden patterns or structures, negating the need for labelled data, and anomaly detection (Usama et al, 2019). Weaknesses include challenges with interpretability and performance measurement, and algorithm sensitivity to initial setup or hyperparameters (e.g., the number of clusters in K-means clustering).

In summary, choosing the most suitable algorithm depends on the problem, nature of available data, and goal of the analysis.

References:

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A. and Aljaaf, A.J., 2020. A systematic review on supervised and unsupervised machine learning algorithms for data science. Supervised and unsupervised learning for data science.

Bell, J. (2020) Machine Learning: Hands-On for Developers and Technical Professionals. Wiley.

Huang, T., Chen, K., Li, B., Liu, Y.H. and Dou, Q. (2023). Guided reinforcement learning with efficient exploration for task automation of surgical robot. arXiv preprint arXiv:2302.09772.

Qayyum, A., Qadir, J., Bilal, M. and Al-Fuqaha, A. (2020). Secure and robust machine learning for healthcare: A survey. IEEE Reviews in Biomedical Engineering, 14.

Russell, S. and Norvig, P. (2021) Artificial Intelligence: A Modern Approach, Global Edition. 4th ed. Pearson Education Limited.

Singh, A., Thakur, N. and Sharma, A. (2016). A review of supervised machine learning algorithms. In 2016 3rd international conference on computing for sustainable global development (INDIACom) IEEE.

Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K.L.A., Elkhatib, Y., Hussain, A. and Al-Fuqaha, A. (2019). Unsupervised machine learning for networking: Techniques, applications and research challenges. IEEE access.

Yu, K.H., Beam, A.L. and Kohane, I.S. (2018). Artificial intelligence in healthcare. Nature biomedical engineering, 2(10).