University of Essex
MSc Artificial Intelligence
Module: UAI_PCOM7E January 2025 A
Unit 8: Collaborative Discussion – Business Use Cases for AI

Initial Discussion

As a cloud developer, one facet of my work involves building out back end systems for detecting and remediating cases of copyright infringement of the company's social media assets; for example, video clips shared on social media platforms by fans of the company's movies and TV series. Some instances of media sharing are approved and encouraged by the company, such as promotional trailers. Others are not approved and subject to copyright take down actions.

Novel cases of infringement have started to crop where fans of protected content use generative AI platforms to modify the source video tracks, most often to generate fan works based on the original content, such as by extending scenes or changing the characters in scenes. This material can be more difficult to detect than straightforward instances of piracy, since the original content was heavily modified and stripped of its identifying metadata. Depending on the type and extent of the transformation of the original material, however, it still can be subject to copyright protection.

As such, a novel use for AI systems would be to automate detection and reporting of this type of unsanctioned modification of protected intellectual property, rather than relying on human employees to find and act on it. The challenge, however, is that generative AI platforms are rapidly changing, with new ones released to the public regularly, which makes defining a usable programmatic taxonomy for markers of infringement in such video files difficult (Lin et al., 2024).

Reference

Lin, L., Gupta, N., Zhang, Y., Ren, H., Liu, C.-H., Ding, F., Wang, X., Li, X., Verdoliva, L. and Hu, S. (2024). *Detecting Multimedia Generated by Large AI Models: A Survey*. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.2402.00045.

Peer Response No. 1

Although the linear regression is not the most commonly used among all supervised learning algorithm types, but it has its own uses. As it's mostly used for discovering the relation between a certain data point, prediction, forecasting and etc. it might work by plotting as the x-axis will be independent variable while y-axis is a dependent variable. Accordingly, will be plotted out in linear fashion to determine the expected future data. No doubt such method might be very helpful to predict the future expenses of any company or the expected performance, and such prediction might be very helpful to tackle a big obstacle which might be unexpected without the linear regression method. The linear regression it really gets advantage of gathered data by using them as independent variable while the expected data will the dependent variable and by having both of the variable, we can draw a linear which represent the expected future values, and it's considered as really simple method in the implantation. Obviously the linear is not comprehensive method to calculate everything in one method but I believe it's really important category of the supervised learning algorithms.

References (Provided But Not Cited)

Tableau (2025). Artificial intelligence (AI) algorithms: a complete overview. *Tableau*. Available at: https://www.tableau.com/data-insights/ai/algorithms.

Gupta, M. (2018). ML | Linear Regression – GeeksforGeeks. *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/ml-linear-regression/.

Peer Response No. 2

Linear regression and logistic regression are indeed two fundamental ML algorithms. Moreover, they are also basic algorithms in statistics, from which their ML versions originate. I will elaborate on this a bit further.

Both traditional linear regression and ML linear regression rely on the same mathematical foundations, although they are used for different purposes today. Traditional linear regression remains a tool for explaining relationships between the dependent and independent variables. The regression model is fitted usually by the Ordinary Least Squares (OLS), which is an optimization method that has an analytical solution. It assumes linearity and normality. On the other hand, ML linear regression is primarily used for making predictions. It generally uses some form of gradient descent as an optimization method, is scalable to large datasets and does not impose strict assumptions.

In the case of logistic regression the situation is quite similar. The traditional version focuses on explaining relationships and uses Maximum Likelihood Estimation (MLE), while the ML version is used as a predictive model that minimizes classification error using some gradient-based method.

In conclusion, the ML versions of both algorithms are adaptations of their traditional counterparts, optimized for prediction and scalability rather than hypothesis testing and inference.

I found your post informative, with its strength being the presentation of both the advantages and potential risks. I wanted to add some background information about the evolvement of these algorithms.

References (Provided But Not Cited)

Montgomery, D. C., Peck, E. A. and Vining, G. G. (2021). *Introduction to Linear Regression Analysis (6th ed.)*. Wiley.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Peer Response No. 3

This post excellently covers the linear and logistic regression concepts as basic algorithms in the domain of supervised machine learning. Still, it overlooks some important evaluative shortcomings by glossing the intricacies and intricacies and risks of these techniques. For example, while the post discusses the overfitting and feature misidentification problems, there is little discussion about overarching societal repercussions of such model implementation, especially in sensitive domains such as healthcare or criminal justice. An AI trained on incomplete and biased data sets has the ability to reinforce existing disparities, which is an unjust outcome (Mehrabi et al, 2021).

Furthermore, the SOC analysis of the impacts of AIs is less thorough as it overlooks other important scoped risks such as data leaks and adversary attacks, and model's interpretability. Data leaks, for instance, can lead to the prediction being classed as an "accurate" prediction, while they are not in fact classed as such (Goodfellow, Shlens and Szegedy, 2015). In addition, the opaque nature of the decision-making process in these models breeds distrust, especially in situations where trusting these models is crucial (Rudin, 2019).

Although the post provides a good insight into the details of the technical aspects of linear and logistic regression, it should broaden its scope to tackle the ethical and pragmatic boundaries. As these issues have to be resolved, the equity, trustworthiness, and responsibility of AI systems in practical uses can be guaranteed and preserved.

References

Goodfellow, I. Shlens, J., and Szegedy, C. (2015). Explaining and harnessing adversarial examples. [arXiv preprint] arXiv:1412.6572.

Mehrabi, N, Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, vol. 54, no. 6: 1-35.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, vol. 1, no. 5: 206-215.

Peer Response No. 4

The post discusses the important algorithms of supervised machine learning which include linear and logistic regression. The practical implementations through your examples demonstrate how linear regression measures rainfall while logistic regression identifies image classifications.

I value your notice about how simple it is to implement these models which makes them practical.

The major problems concerning model overfitting together with achieving good generalization performance on new data remain vital aspects you have correctly identified. Introducing Ridge or Lasso regularization techniques will strengthen linear regression applications by adding penalty terms to prevent coefficient swelling in unpredictable situations. The classification issues of logistic regression could be solved through non-linear techniques and ensemble methods such as Random Forests when analyzing complex cases that require non-linear classification boundaries. You raise important questions about ethical aspects when deploying AI systems. As AI specialists, we have an important role to play, we need to execute model development that performs with fairness and transparency so users obtain clear understandings about the decision-making process.

Your post provides significant contributions that expand our knowledge about these algorithms both in strengths and weaknesses.

References

Not provided.

Discussion Summary

The discussion revolved around comparing and contrasting the fundamental characteristics of linear and logistic regression models in supervised machine learning.

Linear regression models were broadly defined as designed to make predictions consistent with a linear relationship between input values and predicted values (Burkov, 2019); for example, the predicted level of rainfall given an input date value as plotted on a two dimensional graph.

Peer replies noted that linear regression have the advantage of relative simplicity to implement and interpret (Burkov, 2019; Hewamalage, Ackermann, and Bergmeir, 2022; Montgomery, D., Peck, E. and Vining, G., 2021). A common pitfall stems from the model overfitting data, or making erroneous correlative predictions between input and output values (Burkov, 2019; Hewamalage, Ackermann, and Bergmeir, 2022; Montgomery, D., Peck, E. and Vining, G., 2021).

Logistic regression models were defined as intended to make binary categorization decisions, such as identifying whether or not a data point does or does not belong in an identity category established by the training data set (Burkov, 2019); for example, whether or not a given photo does or does not depict a vehicle in the “car” category. Similarly, peer replies noted that logistic regression models are also relatively simple to implement and interpret (Burkov, 2019; Hewamalage, Ackermann, and Bergmeir, 2022; Montgomery, D., Peck, E. and Vining, G., 2021). Logistic regression models, however, are prone to defining category membership based on irrelevant or nonprobative details (Hewamalage, Ackermann, and Bergmeir, 2022); for example, assigning a sample image to a “car” category because of irrelevant details like image background noise.

Peer replies also noted that significant real world consequences may result from of both types of models reaching erroneous predictions that are uncritically relied upon by decision makers (Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A., 2021), such that thoroughly testing the probative reliability of any models used is a practical and ethical necessity.

References

Burkov, A. (2019). *The Hundred-Page Machine Learning Book*. Andriy Burkov.
Goodfellow, I., Shlens, J. and Szegedy, C. (2014). *Explaining and Harnessing Adversarial Examples*. [online] arXiv.org. Available at: https://arxiv.org/abs/1412.6572.

Hewamalage, H., Ackermann, K. and Bergmeir, C. (2022). Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*. doi:https://doi.org/10.1007/s10618-022-00894-5.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, [online] 54(6), pp.1–35. doi:https://doi.org/10.1145/3457607.

Montgomery, D., Peck, E. and Vining, G. (2021). *Introduction to Linear Regression Analysis*. [online] *Google Books*. John Wiley & Sons. Available at: https://books.google.com/books?hl=en&lr=&id=tCIgEAAAQBAJ&oi=fnd&pg=PR13&dq=Montgomery.