

How Statistics Supports Modern Machine Learning

Machine learning and data science depend heavily on statistics. Statistics helps computers learn patterns from data choose the most useful information and check whether models make reliable predictions. This report explains how statistical ideas are used when building and evaluating machine learning models.

Statistical Foundations of Machine Learning

Statistics provides the basic ideas that machine learning can be seen as a way of using statistical models to find patterns and relationships in data. These models help estimate how different variables are related and allow computers to make predictions based on past data. For example, methods like linear regression and logistic regression come directly from statistics. These techniques estimate model parameters so that the model best fits the observed data. which happens when a model learns the training data too closely and performs poorly on new data.

Feature Selection and Statistical Thinking

Statistics also helps decide which variables or features should be used in a model. In large datasets not every variable is useful. Statistical techniques such as correlation analysis hypothesis testing, and regularisation help identify the most important features. Research by Guyon and Elisseeff shows that feature selection improves machine learning models by removing irrelevant or noisy variables. This makes models simpler faster and often more accurate. Statistical thinking also helps researchers understand whether relationships in the data are meaningful or just random.

Model Validation and Evaluation

Another important use of statistics in machine learning is evaluating model performance. Techniques such as cross validation allow researchers to test how well a model works on new, unseen data. Statistical tools also help measure uncertainty and determine whether a models results are reliable.

Conclusion

In summary, statistics plays a key role in modern machine learning. It supports model development, helps select important features, and provides reliable ways to evaluate model performance. Without statistical methods, machine learning models would be less accurate, harder to interpret, and less trustworthy.

References

- [1] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York: Springer, 2009.
- [2] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press, 2012.