e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

MODEL SELECTION IN MACHINE LEARNING

Shilpa Mathur*1

*1Thakur College Of Engineering & Technology, Mumbai, India.

DOI : https://www.doi.org/10.56726/IRJMETS32850

ABSTRACT

There may be situations wherein a model performs well on training data but not on the test data. One can also face confusion regarding the model to be used for a given problem. For example, by now, you have learnt about many classification models. Given a problem that requires classification, how would you decide which model is the best? Questions such as these frequently arise irrespective of the choice of the model, data or the problem itself. This paper discusses about Model selection criteria in Machine Learning.

I. INTRODUCTION

The central issue in all of Machine Learning is "how do we extrapolate what has been learnt from a finite amount of data to all possible inputs 'of the same kind'?". We build models from some training data. However the training data is always finite. On the other hand the model is expected to have learnt 'enough' about the entire domain from where the data points can possibly come. Clearly in almost all realistic scenarios the domain is infinitely large. How do we ensure our model is as good as we think it is based on its performance on the training data, even when we apply it on the infinitely many data points that the model has never 'seen' (been trained on)?

Occam's Razor is a predictive model has to be as simple as possible, but no simpler. Often referred to as the Occam's Razor, this is not just a convenience but a fundamental tenet of all of machine learning.

To measure the simplicity, we often use its complementary notion — that of the complexity of a model. More complex the model, less simple it is. There is no universal definition for the complexity of a model used in machine learning.

Model Selection Criteria

However here are a few typical ways of looking the complexity of a model.

1. Number of parameters required to specify the model completely. For example in a simple linear regression for the response attribute y on the explanatory attributes x1,x2,x3 the model y = ax1 +bx2 is 'simpler' than the model y = ax1 +bx2 +cx3 — the latter requires 3 parameters compared to the 2 required for the first model.

2. The degree of the function, if it is a polynomial. Considering regression again, the model y = ax2 1 +bx3 2 would be a more complex model because it is a polynomial of degree 3.

3. Size of the best-possible representation of the model. For instance the number of bits in a binary encoding of the model. For instance more complex (messy, too many bits of precision, large numbers, etc.) the coefficients in the model, more complex it is. For example the expression (0.552984567 ∗ x 2 + 932.4710001276) could be considered to be more 'complex' than say (2x +3x 2 +1), though the latter has more terms in it.

4. The depth or size of a decision tree. Intuitively more complex the model, more 'assumptions' it entails. Occam's Razor is therefore a simple thumb rule — given two models that show similar 'performance' in the finite training or test data, we should pick the one that makes fewer assumptions about the data that is yet to be seen. That essentially means we need to pick the 'simpler' of the two models.

In general, among the 'best performing' models on the available data, we pick the one that makes fewest assumptions, equivalently the simplest among them. There is a rather deep relationship between the complexity of a model and its usefulness in a learning context. We elaborate on this relationship below.

1. Simpler models are usually more 'generic' and are more widely applicable (are generalizable). One who understands a few basic principles of a subject (simple model) well, is better equipped to solve any new unfamiliar problem than someone who has memorized an entire 'guidebook' with a number of solved examples (complex model). The latter student may be able to solve any problem extremely quickly as long as it looks similar to one of the solved problems in the guidebook. However given a new unfamiliar problem that doesn't fall neatly into any of the 'templates' in the guidebook, the second student would be hard pressed to solve it

than the one who understands the basic concepts well and is able to work his/her way up from first principles.

• A model that is able to accurately 'predict

2. Simpler models require fewer training samples for effective training than the more complex ones and are consequently easier to train. In machine learning jargon, the sample complexity is lower for simpler models.

3. Simpler models are more robust — they are not as sensitive to the specifics of the training data set as their more complex counterparts are. Clearly we are learning a 'concept' using a model and not really the training data itself. So ideally the model must be immune to the specifics of the training data provided and rather somehow pick out the essential characteristics of the phenomenon that is invariant across any training data set for the problem. So it is generally better for a model to be not too sensitive to the specifics of the data set on which it has been trained. Complex models tend to change wildly with changes in the training data set. Again using the machine learning jargon simple models have low variance, high bias and complex models have low bias, high variance. Here 'variance' refers to the variance in the model and 'bias' is the deviation from the expected, ideal behaviour. This phenomenon is often referred to as the bias-variance trade-off.

4. Simpler models make more errors in the training set — that's the price one pays for greater predictability. Complex models lead to overfitting — they work very well for the training samples, fail miserably when applied to other test samples.

II. CONCLUSION

The validity of Occam's razor has long been debated. Critics of the principle argue that it prioritizes simplicity over accuracy and that, since one cannot absolutely define “simplicity,” it cannot serve as a sure basis of comparison.

III. REFERENCES

[1] https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML

[2] The Hundred-Page Machine Learning Book by Andriy Burkov.

[3] Machine Learning For Absolute Beginners by Oliver Theobald.

[4] Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow by Geron Aurelien.

[5] Machine Learning for Hackers by Drew Conway and John Myles White.

[6] Fundamentals of Machine Learning for Predictive Data Analytics by John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy.