



Technical Report No. 145

Popper, Falsification and the VC-dimension

David Corfield,¹ Bernhard Schölkopf,¹
Vladimir Vapnik²

November 2005

¹ Department of Empirical Inference, email: dcorfield;bs@tuebingen.mpg.de,

² NEC Laboratories, 4 Independence Way, Princeton, NJ 08540, USA, email: vapnik@att.net

Popper, Falsification and the VC-dimension

David Corfield, Bernhard Schölkopf, Vladimir Vapnik

Abstract. We compare Sir Karl Popper's ideas concerning the falsifiability of a theory with similar notions from VC-theory. Having located some divergences, we discuss how best to view Popper's work from the perspective of statistical learning theory.

1 Introduction

In [Va95] the philosopher Sir Karl Popper's treatment of falsificationism is mentioned as an important inspiration for statistical learning theory. In his 1934 book *Logik der Forschung*, first translated into English in 1959, Popper marked out his differences from members of the *Vienna Circle* by maintaining that the single factor distinguishing science from other claims to knowledge was its preparedness to risk falsification. Taking Albert Einstein as a paradigmatic scientist, Popper was impressed by Einstein's willingness to make bold conjectures with testable predictions. Starlight would be found to bend as it travelled to our telescopes past the sun, and bend by a specified amount. If, within experimental error, this precise bending was not found to have taken place, Einstein would be prepared to give up on his general theory of relativity, given that he was assured that the observations had been conducted properly.

On the other hand, if the right amount of bending was observed, this would not *confirm* the theory. What was at stake for Popper was simply a point of logic: universal statements (with infinite domain) can be falsified, but cannot be verified. Popper could make no sense of the idea that a scientific theory becomes more probable when a prediction is verified. Rather, all we can say is that the theory has been severely tested, or well-corroborated. The Vienna Circle, and associated logical empiricists, meanwhile, did not agree with this asymmetric treatment of verification and falsification, and developed a theory of confirmation in a probabilistic framework. Much of *The Logic of Scientific Discovery* is taken up with a critique of the use of probabilistic representations of states of knowledge. Popper has thus become a rallying point for anti-Bayesian philosophers of science.

In this technical report, we shall examine Popper's ideas in the light of statistical learning theory. We shall scrutinise divergences we have found between Popper's idea of falsifiability and the VC-dimension, and explain how Popper might have viewed the divergence.

2 Falsification and the VC-Dimension

Recall that the VC (Vapnik-Chervonenkis)-dimension [Va95] of a set of classifying hypotheses is the largest natural number, n , such that there is a set of n points, for which however they are labelled '+' or '-', there is a hypothesis which agrees with this labelling. The set of points is said to be shattered by the hypothesis class. For example, the set of half-planes has VC-dimension three because any set of three non-collinear points is shattered by these classifiers, while any set of four points is not shatterable by the same class. (Think of the corners of a square, opposite ends of each diagonal with matched labelling.) One dimension higher, the set of half-spaces has VC-dimension four.

The VC-dimension is used in the following kind of way. Imagine measurements are taken of the weight, height and waist size of a large sample of people from a given population. Someone proposes that there is a linear classifier (a separating plane) which can accurately separate the males from the females within this population, using the above measurements as inputs. The VC-dimension of this class of classifiers is four. If the data is plotted in three-dimensional space and four points are selected which do not lie on the same plane, then whatever the genders of the people corresponding to these points, a plane could be found to separate the males from the females. Now, if we take our large sample and find an accurate linear classifier, then VC-theory gives us results of the form: with probability $x\%$ the classifier will be $y\%$ accurate on future observations if drawn independently from the same distribution, where y depends not only on x , but also on the sample size and VC-dimension.

This idea of a set of points being shattered by a class of hypotheses may bring to mind Karl Popper's notion of non-falsifiability in the following sense. You cannot falsify the theory that a linear classifier can separate the

genders by looking at a sample of size four however they are labelled. It will pay us then to look more closely at Popper's writings to see if he says anything along these lines.

3 The Logic of Scientific Discovery

In The Logic of Scientific Discovery Popper describes two methods for comparing theories:

1. The containment relation between classes of falsifiers

This is determined by the degree of universality and the degree of precision (of predicate and of measurement) of the theory. So "all planets move in ellipses" is less universal and less precise than "all heavenly bodies move in circles". The collection of falsifiers of the first theory is contained within the collection of falsifiers of the second theory, so the latter is more falsifiable.

2. The dimension of a theory

If there exists, for a theory t , a field of singular (but not necessarily basic) statements such that, for some number d , the theory cannot be falsified by any d -tuple of the field, although it can be falsified by certain $(d + 1)$ -tuples, then we call d the characteristic number of the theory with respect to that field. All statements of the field whose degree of composition is less than d , or equal to d , are then compatible with the theory, and permitted by it, irrespective of their content. [Po59, 130]

At first glance Popper's dimension of a theory seems to be a precursor of the VC-dimension. But let us look more closely. The point about statements being "singular (but not necessarily basic)" is intended in the sense that a recording of the position of an event in d -dimensional Euclidean space would count as one statement, rather than as d statements concerning its co-ordinates. This agrees with the VC-dimension set-up, so let us calculate the dimension of the theory which states that labelled points in the plane can be separated by a line. Clearly any two points can be separated. However, there are configurations of three points and labellings which cannot be separated, for example, three collinear points labelled '+', '-', '+'. The existence of this non-generic configuration means that the Popper dimension of the theory is two. And indeed the Popper dimension of hyperplane classifiers in any d -dimensional space is two, where the VC-dimension can be shown to be $(d + 1)$.

4 What to make of Popper

There are two ways to respond to this discrepancy we have observed:

1. First, we might say that Popper simply made a mistake.
2. Second, we might say that Popper is simply not interested in the same problems as Vapnik and Chervonenkis.

4.1

One could argue that Popper was just being imprecise in his definition. If he had insisted that points could not be specified with complete accuracy, then one could not observe perfectly collinear points, and the Popper-dimension for separating half-spaces would equal the VC-dimension.¹ However, Popper's treatment of a different situation suggests he had not thought through his definition. We find that he gives the dimension of the theory that "All planets move in ellipses" as five, although, strictly speaking, the observation of three collinear points on the orbit would falsify the theory. Allowing for a margin of error in the observations, however, three points could be accommodated.² But still, very generous margins of error would be necessary to allow some configurations of four points to be situated on an ellipse. So strictly speaking, the Popper-dimension of this theory is two, and disallowing perfect precision of observations it is three.³

This relates to a more general mistake concerning his treatment of simplicity and dimension. Popper claims: "In algebraic representation, the dimension of a set of curves depends upon the number of parameters whose values we may freely choose. We can therefore say that the number of freely determinable parameters of a set of curves by which a theory is represented is characteristic for the degree of falsifiability (or testability) of that theory." [Po59,

¹On the other hand, these dimensions would diverge if a minimum margin were required.

²Even here, though, configurations of three very nearly collinear points might require an ellipse with a very large major axis relative to the distances between the points.

³Peter Turney [Tu91] has also located this confusion on Popper's part.

131]. So he might identify simplicity either with degree of falsifiability or number of parameters. But he comes down in the end in favour of the former:

The epistemological questions which arise in connection with the concept of simplicity can all be answered if we equate this concept with degree of falsifiability. [Po59, 140]

Above all, our theory explains why simplicity is so highly desirable. To understand this there is no need for us to assume a 'principle of economy of thought' or anything of the kind. Simple statements, if knowledge is our object, are to be prized more highly than the less simple ones because they tell us more; because their empirical content is greater; and because they are better testable. [Po59, 142]

So he appears to have decided in favour of the degree of falsifiability as to the strength of a theory, although confusingly he equates this to simplicity. However, he then continues to describe sine oscillations as simple [Po59, 143], giving no thought to the number of points needed to falsify this class.⁴ The use of the VC-dimension in statistical learning theory runs against the idea that the generalisability of a theory goes along with its simplicity, calculated in terms of the number of its parameters. Standard examples exist of hypothesis classes with low VC-dimension and a very large number of parameters (e.g., support vector machines), and *vice versa*. The sine curves Popper saw as easily falsifiable are often given as a paradigmatic example of the latter case. The set $\{sign(\sin bx) : b > 0\}$ can shatter any finite set of points distributed along the x axis.

4.2

Let us turn now to the more charitable reading of Popper - that his concerns are not those of statistical learning theory. Where statistical learning theory operates in situations where the data we receive is recorded passively, with a view to giving us confidence that we can expect classification or regression to continue to work accurately if further incoming data is independent and from the same distribution, Popper is interested in the situation confronting the bold research scientist, who will leave no stone unturned in their quest for truth:

...our methodological decision - sometimes metaphysically interpreted as the principle of causality - is to leave nothing unexplained, i.e. always to try to deduce statements from others of higher universality. This decision is derived from the demand for the highest attainable degree of universality and precision, and it can be reduced to the demand, or rule, that preference should be given to those theories which can be most severely tested. [Po59, 123]

Rather than happily anticipating future accurate predictions on new data, Popperian scientists go out of their way to find the weaknesses of their theory. In other words, Popper appears to be more interested in what is called active learning, and indeed the active learning of generative models. This isn't about settling for a classifier for which we have probabilistic guarantees that its future error rate will be less than some figure. Rather, theories are there to be shot down. Let's illustrate this point with an example.

Returning to our example of the weight, height and waist size of a population, we mentioned above that the VC-dimension of the class of linear classifiers is four. The Popper-dimension for this class by contrast is only two according to his definition if infinite accuracy is allowed. Where VC-theory gave results about our expectation of an accurate classifier continuing to perform well on unseen data, Popperian scientists dealing with the proposed hypothesis are looking to falsify it. Having found a woman with measurements (60kg, 160cm, 80cm) and a man with measurements (70kg, 170cm, 90cm), if they can find a woman whose measurements are, for instance, (80kg, 180cm, 100cm), they will have succeeded in falsifying the hypothesis.

Reading Popper as interested in active learning makes sense of a further discrepancy between Popper and VC-theory. Recall that Popper's first criterion for comparative falsifiability states that if the falsifiers of one theory are contained within those of another, the latter is more falsifiable. Increasing the domain of a theory allows further opportunities for falsification to occur. In VC-theory, on the other hand, when the domain of a set of functions is increased, its VC-dimension can only increase, hence the set is less falsifiable.⁵

A neat way of summarising the situation, then, might be to congratulate Popper for his original idea of falsifiability, then to say that it acts as a *motif*, appearing in different guises in the kind of learning that is science and the

⁴One might try to save Popper again by requiring a minimum margin to a sinusoidal classifier, or to limit the precision with which a parameter could be specified.

⁵This difference is made less decisive if one considers the VC-entropy, rather than the VC-dimension. Since the VC-entropy depends on the distribution, it is not possible to say in general how it is affected by changes of the domain.

kind that is machine learning. Popper saw science as a quest for truth, its theories not subject to static appraisal. As we mentioned in the introduction, Popper was strongly opposed to any notion of confirmation:

Like inductive logic in general, the theory of the probability of hypotheses seems to have arisen through a confusion of psychological with logical questions. Admittedly, our subjective feelings of conviction are of different intensities, and the degree of confidence with which we await the fulfilment of a prediction and the further corroboration of a hypothesis is likely to depend, among other things, upon the way in which this hypothesis has stood up to tests so far - upon its past corroboration. But that these psychological questions do not belong to epistemology or methodology is pretty well acknowledged even by believers in probability logic. (Note: I am alluding here to the school of Reichenbach rather than to Keynes.) [Po59, 255]

And this isn't to make a distinction between the probable truth of a hypothesis and its probable accuracy. Nowhere in [Po59] does he say anything about how reliable are our well-corroborated theories as we travel the path towards truth, i.e., how much we can trust their predictions. Now this is rather peculiar. Surely we rely on what our scientists tell us are our best confirmed theories. For example, we believe their predictions of the next solar eclipse. This isn't just blind faith. We seem here to be treating scientific theories as we might the hypotheses in the output of a machine learning algorithm, i.e., as reliable predictors. Where does the confidence arise from?

In 1963 Popper introduced the concept of *verisimilitude* in his book *Conjectures and Refutations* [Po63]. With this concept he was attempting to measure the truthlikeness of a theory by taking the difference between its truth content and its falsity content. We need not enter into details here as this idea has almost universally been seen to have failed, but it does point to an interest on his part in relations between the successes and failures of a theory. In view of the fact that statistical learning theory has developed such a relation in the form of the *generalization error*, we might speculate that something akin to this could have filled the lacuna in Popper's theories in a way which might have satisfied him, avoiding as it does the notion of the probable truth of a theory.

5 Conclusion

It seems clear that Popper can only be rescued from the charge that he made serious mistakes in *The Logic of Scientific Discovery* by generous interpretation, evidence that he had any inkling of the need for which is lacking. However, we do not think it right merely to say that Popper had a partial insight into an important component of statistical learning theory. His first concern was with a different kind of learning, that conducted by the 'bold and daring' research scientist. This having been said, however, there appears to be no way for Popper to speak about the reliability of well-tested theories, and yet one surely needs to be able to speak of one's confidence in the future predictions of such theories. We suggest that statistical learning theory might have satisfied him on this score.

References

- [Po59] Karl Popper, *The Logic of Scientific Discovery*, Hutchinson, translation of *Logik der Forschung*, 1994.
- [Po63] Karl Popper, *Conjectures and Refutations*, Routledge, 1963.
- [Tu91] Peter Turney, 'A Note on Popper's Equation of Simplicity with Falsifiability', *British Journal for the Philosophy of Science* **42**: 105-109, 1991.
- [Va95] Vladimir Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.