TABLE OF CONTENTS:

Term Paper 2

1 Motivation 2

2 Aim 2

3 Introduction 2

4 Popper’s Falsification and Dimension of a theory 3

4.1 The Containment Relation between Classes of Falsifiers 3

4.2 The Dimension of a theory 3

5 Comparison to VC Dimension 4

6 Popper’s notion of simplicity 5

7 Critique of Karl Popper’s ideas 6

8 What do we learn? 6

9 References 7

Term Paper

Philosophical Motivation for VC Dimension
ELL-880

Harsh Maheshwari - 2015EE30517

1 Motivation

Learning theory uses statistics to answer questions which were discussed in philosophy for a long time. For example, Epistemology, a branch in philosophy, deals with study of the nature of knowledge, justification, and the rationality of belief. All these themes are linked closely to Learning Theory and thus it would be interesting to see philosopher's view on such topics, how did they deal with it?, What can we learn?

2 Aim

In the book, 'The Nature of Statistical Learning'[1] - V.P. Vapnik, Philosopher Sir Karl Popper's treatment of falsification has been considered an important inspiration of statistical learning theory and VC Dimension. In this term paper, we will look at some of Karl Popper's ideas, see their links to Learning theory and compare them with Statistical Learning Theory's ideas. We will also try to understand what Popper had in mind, criticism to his ideas and what can we learn from those.

3 Introduction

Over the past decade and a half, a paradigm shift has occurred in the field of machine learning. Data with possibly many thousands of attributes, such as the values of the pixels of a digital photograph, can be handled by powerful classifiers to allow accurate labelling. Rather than forming a model to represent the data, this style of machine learning aims simply to be able to discriminate between inputs in order to label them correctly.
The theoretical basis for these discriminative classifiers is known as statistical learning theory. Here, as elsewhere in inductive learning, there is an important balance to be struck between accuracy and overfitting. Overfitting occurs when too rich a space of hypotheses is used to represent a data set. Now, in statistical learning theory the richness of the hypothesis space is not controlled by the degree of a curve or the number of parameters of a hypothesis, but by a construction known as the VC-dimension. This dimension, as we shall see, can be thought of as measuring a degree of falsifiability. It is not surprising then that Karl Popper's name crops up in discussions of statistical learning theory. Karl Popper's dimension of a theory relates closely to the VC Dimension, Interestingly, V.P. Vapik, in his book has considered his ideas as important

inspiration of statistical learning theory and VC Dimension. Let us look at Sir Karl Popper's ideas of falsification and dimensionality of a theory.

4 Popper's Falsification and Dimension of a theory

Much of Popper's early work in the philosophy of science focuses on what he calls the problem of demarcation, or the problem of distinguishing scientific (or empirical) theories from non-scientific theories. In particular, Popper aims to capture the logical or methodological differences between scientific disciplines, such as physics, and non-scientific disciplines, such as myth-making.
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, even if the right amount of bending was observed, this would not *confirm* his theory.

What Popper had in mind was simply the point of logic that synthetic universal statements which have infinite domains can never be verified but only be falsified. In simpler terms, just like in mathematics, we cannot prove a claim or theorem by showing validity on some examples but to disprove, only one contradiction is enough. Popper thus could make no sense of the idea that a scientific theory (again with infinite domain) would become more probable of being true once a prediction is verified.

In development of this conception of scientific learning, he developed a way to compare and quantify theories as to their riskiness, on their potential to be falsified.

4.1 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]

4.2 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. [2]

At first glance, Popper's Dimension of a theory (PDoT) may look very similar to VC Dimension. But there is a subtle difference in the two definition which we will find out in the next section.

5 Comparison to VC Dimension

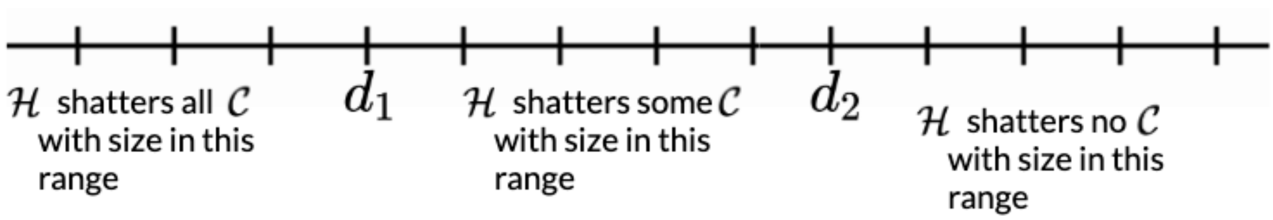
First, let us remind ourselves of VC Dimension.

The VC-dimension of a set of classifying hypotheses is the largest natural number, n, such that there is a set of n distinct points, for which, however, they are labelled '+' or '-', there is a hypothesis which agrees with this labelling.

- If VCDim(H) = d
  - ∃ C ⊆ X, |C| = d, H shatters C
  - ∀ C ⊆ X, |C| = d+1, H does not shatter C
- But in similar fashion, if PDoT(H) = d
  - ∀ C ⊆ X, |C| = d, H shatters C (cannot be falsified)
  - ∃ C ⊆ X, |C| = d+1, H does not shatter C

The swap between existentially and universality in the sub points makes these two characteristics of a hypothesis class different. In words, the VC-dimension is the largest number of points one can shatter, the Popper dimension is one less than the smallest number of points one can not shatter.

For an hypothesis class H, we know that if H can shatter all C of size d, it can also shatter all sets of size d-1, we can say that



then we can see that according to the two definitions,
PDoT(H) = d_1
VCDim(H) = d_2

And thus PDoT(H) is a lower bound on VCDim(H), VCDim(H) $\geq$ PDoT(H)

There can be two ways to respond to the discrepancy observed above, one is to believe Popper simply made mistake and wasn't very precise in his definitions. I believe this could be true because at lots of instances in his book he has been imprecise in the math involved, for example, 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.

The other way to respond to this discrepancy is to believe that Popper had a different question in mind than the one Statistical Learning Theory tries to solve. Before coming to a conclusion on this discrepancy, let us see what were Popper's views on Simplicity of a theory and how that connects to Statistical Learning Theory.

6 Popper’s notion of simplicity

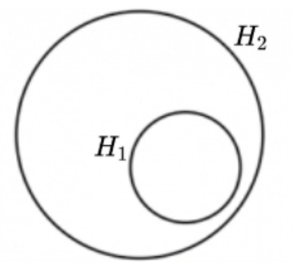
The view that simplicity is a virtue in scientific theories and that, other things being equal, simpler theories should be preferred to more complex ones has been widely advocated in the history of science and philosophy.
It often goes by the name Ockham’s Razor - Simplicity ought to be one of the key criteria for evaluating and choosing between rival theories, alongside criteria such as consistency with the data and coherence with accepted background theories.
There has been a lot of discussions in Philosophy to explain why the thrive for simplicity is important for science. Popper tries to explain it using his theory of falsification.

Popper believes,

> *“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.”[2]*
>
> *“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.”[2]*

Thus, Popper equated simplicity of a theory by its falsifiability (which is inversely related to the PDoT) and unsurprisingly, one of the ways to quantify complexity of a hypothesis class in Statistical Learning Theory is its VC Dimension.

Let H_1 and H_2 be two hypothesis classes such that $H_1 \subset H_2$
and let F_1 be the set of falsifiers of H_1 and F_2 be the set of falsifiers for H_2



For any $f \in F_2$, we can say that none of the hypothesis in H_2 ‘explains’ it, which implies, none of the hypothesis in H_1 can ‘explain’ f

$\implies \forall f \in F_2, f \in F_1$

$\implies F_2 \subset F_1$

And by Popper’s ‘The containment relation between classes of falsifiers’ (Section 4.1) we can say that H_1 is more falsifiable than H_2, thus, according to Popper more ‘simpler’ than H_2 since he equates simplicity with falsifiability.
This also is consistent with Statistical Learning Theory because $VCDim(H_1) < VCDim(H_2)$ and thus H_1 would be more ‘simple’ than H_2

Popper believes science should strive for simplicity because such theories will be more quickly eliminated if they are in fact false. The practice of first considering the simplest theory consistent with the data provides a faster route to scientific progress. Importantly, for Popper, this meant that we should prefer simpler theories because they have a lower probability of being true, since, for any set of data, it is more likely that some complex theory (in Popper’s sense) will be able to accommodate it than a simpler theory. [3]

7 Critique of Karl Popper's ideas

Although there is so much similarities in Sir Karl Popper's ideas ([2], 1934) and Statistical Learning theory (1970s), there has been a lot of critique about his work mainly focusing on the differences between these two.
One such critique ([4], P. Turney) is where the authors bring to notice a flaw in one of his examples for equating simplicity to falsifiability. They point out the fact that Popper had been using *dimension* to mean both Geometrical and Theoretical Dimension and thus fails to acknowledge the fact that these two can be different. They note that a class of circles and a class of ellipses have the same theoretical dimension (falsifiability) but they don't have the same 'simplicity'.

Corfield et al.[5] also point out inconsistencies within Popper's book like many time assigning the number of parameters as the dimension of theory without thinking that they may not be the same. These inconsistencies have called a lot of critique of Popper.

Apart from these, Popper's equation of simplicity with falsifiability suffers from some serious objections. First, it cannot be applied to comparisons between theories that make equally precise claims, such as a comparison between a *specific* parabolic hypothesis and a specific linear hypothesis, both of which specify precise values for their parameters and can be falsified by only one data point. It also cannot be applied when we compare theories that make probabilistic claims about the world, since probabilistic statements are not strictly falsifiable. In addition, most philosophers of science now tend to think that falsifiability is not really an intrinsic property of theories themselves, but rather a feature of how scientists are disposed to behave towards their theories. [3]

There have been other criticism too, to Karl Popper's work. The Vienna Circle, and associated logical empiricists, meanwhile, did not agree with this asymmetric treatment of verification and falsification. 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. [5]

8 What do we learn?

Although Popper's work has faced criticism and questioning, I believe what he had in mind was very similar to what Statistical Learning Theory and VC Dimension states, but he was just imprecise in his definitions and examples. Popper's work need a very generous and charitable interpretation and high margin of error (especially for parts where he takes the help of Mathematics to explain his ideas). I do not believe that Popper had only a partial insight to an important part of Statistical Learning Theory but he had attacked the issues right on point and had novel insights and ideas like falsifiability and simplicity which are very similar to what have been incorporated in Statistical Learning Theory.

We know that V.P. Vapnik has written in his book that Sir Karl Popper's ideas were important inspiration to formulate concepts of Statistical Learning Theory. In one way, concepts like VC Dimension are refined and better versions of Popper's ideas. Now these ideas are making their way back to Philosophy. Popper essentially tried to link the simplicity of a theory with the degree to which it can accommodate potential future data: simpler theories are less accommodating than more complex ones. One interesting recent attempt to make sense of this notion of accommodation is due to Harman and Kulkarni (2007). Harman and Kulkarni analyse accommodation in terms of VC Dimension. While Harman and Kulkarni do not propose that VC dimension be taken as a general measure of simplicity (in fact, they regard it as an *alternative* to simplicity in some scientific contexts), ideas along these lines might perhaps hold some future promise for testability/accommodation measures of simplicity. [3]

It is also true that Popper was out to solve a different philosophical question, the question of true scientific discovery, of distinguishing true scientific theories from myths. Popper believed that there is a difference between absolute truth and certainty. The absolute truth is objective and same for all, while certainty is subjective. We can be certain about something which might be false and not certain about something which might be true. We all can be certain about some scientific theories but we can never know which of them is true [6]. Popper believed that theories can only be falsified and never verified. We can never know which theories are true.
Popperian scientist, thus, 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. 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. Popper saw science as a quest for truth, its theories not subject to static appraisal. As we mentioned above, Popper was strongly opposed to any notion of confirmation [5].

We can see through this (rare) example, how philosophy can sometimes motivate important mathematical concepts which can then lay the foundations of practical applications. Philosophers have dealt and are dealing with many important questions about some of the very some fundamental aspects of human life and history and examples like these motivates one to further explore philosophical discussions.

9 References

[1] Vapnik, V. N. (2010). The Nature of Statistical Learning Theory. New York, NY: Springer.

[2] Popper, K. (1959). The logic of scientific discovery, Hutchinson, translation of Logik der Forschung, 1934.

[3] The Internet Encyclopaedia of Philosophy, ISSN 2161-0002, https://www.iep.utm.edu/simplici/, 26-04-2019.

[4] Turney, Peter. "A Note on Popper's Equation of Simplicity with Falsifiability." The British Journal for the Philosophy of Science 42, no. 1 (1991): 105-09. http://www.jstor.org/stable/687610.

[5] Corfield, D., Schölkopf, B., & Vapnik, V.(2005). Popper, Falsification and the VC- dimension (145). Tübingen, Germany: Max Planck Institute for Biological Cybernetics.

[6] Sir Karl Popper: An Interesting Philosopher | Reasonable Faith Video Podcast, https://www.youtube.com/watch?v=-gccBmGrIhs&t=167s

[7] Bandyopadhyay, P. S., & Forster, M. R. (2011). Philosophy of statistics (Vol. 7). Amsterdam: Elsevier, NH.

[8] Stanford Encyclopedia of Philosophy. (n.d.). Retrieved April 26, 2019, from https://plato.stanford.edu/