Liam Bessell
5/3/20
PHIL 351
Dr. Easwaran

On Machines Acquiring Knowledge

Since the mid 20th century computers have become an ever integral part of society. At first they were used to quickly solve large arithmetic problems. Later, home computers became advanced enough to run complex applications like interactive notepads and internet browsers. Recently with the rise of computing power and machine learning, computers have begun "learning" things that weren't hard-programmed into them. But are these machines actually acquiring knowledge, and is that knowledge the same as human knowledge? To answer this question I'm going to first examine Hume's Problem of Induction. After that I will inspect the Bayesian answer to Hume's conclusion. Finally I will extend the Bayesian verdict into the realm of machines and consider what consequences follow from it.

In *An Enquiry Concerning Human Understanding*, David Hume is considering whether or not inductive reasoning is rational. He begins by stating that there are only two forms of arguments: demonstrative and probable. A demonstrative argument is said to follow deductive reasoning, such as if *p* then *q*, *p*, therefore *q*. On the contrary, a probable argument uses inductive reasoning. Consider the probable inference *I* to be of the form all instances of *p* have been *q*, thus the next *p* will also be *q*. Next, Hume claims that all inferences of type *I* rely on the uniformity principle (UP), which is that the future will resemble the past. Hume goes on to claim that since the negation of the UP is not a contradiction it is not demonstrative. He then shows us that any probable argument for UP presupposes UP as one of its premises, thus committing the logical fallacy *Begging the Question*. This is clear by the following argument: The future so far has resembled the past, The future will resemble the past, Therefore the future will resemble the past. He asserts that if there is no UP, then there is no valid chain of reasoning from premises to a conclusion that presupposes UP. Thus, it is not rational to make a probable inference *I*.

In *Bayesianism I* and *II*, Kenny Easwaran explores the fundamental Bayesian stance and cites a Bayesian answer to Hume's Problem of Induction. The Bayesians begin by denying Hume's strong definition of a probable inference. That is, they don't abide by the all-or-nothing approach to beliefs from probable inferences, instead they claim these inferences give us degrees of belief. To illustrate this, Easwaran first considers some universal hypothesis *H* with a non-zero prior. In other words, the a priori probability of *H* being true is greater than zero. For illustration purposes, let *H* be that all swans are white. Easwaran moves on to consider a sequence of observations [E_1, E_2, …, E_n] that are logically entailed by *H*. In other words, each instance of *E* is an observation of a swan, and that observation is entailed by the hypothesis that all swans are white. Next, upon learning [E_1, E_2, …,E_n], conditionalization and Bayes' Theorem yield $P(H)$ / $P(E_1\&E_2\&...\&E_n)$. Now, $P(E_1\&E_2\&...\&E_n) = P(E_1)P(E_2 \mid E_1)...P(E_n \mid E_1\&E_2\&...\&E_{n-1})$ due to the chain rule of probability. Since it follows that the posterior of *H*, $P(H \mid E)$, will eventually go above 1, we have reached a contradiction in the axiom of probability that all probabilities must be greater than or equal to 0 and less than or equal to 1. This must mean that the posterior of *H* converges to 1, rather than going above 1. From this, Easwaran concludes that so long as someone has a non-zero prior in *H*, enough instances of *H* will eventually yield arbitrarily large support for it.

Now that we understand Hume's Problem of Induction and have examined the Bayesian answer to it, we can move forward to consider our central question regarding machines and knowledge. I begin my argument by extending the Bayesian conclusion, namely that enough instances of a hypothesis with a non-zero prior gives an agent arbitrarily large support for believing in the hypothesis to a degree. Next, it is clear that since we can draw arbitrarily large support for a hypothesis based on new evidence, we are rational in taking the hypothesis to be true within a degree of belief. In other words, if my hypothesis is that the sun will rise tomorrow, and I have seen the sun rise everyday of my life so far, my use of Bayesian probable inference to conclude that the sun will most likely rise tomorrow is rational. Indeed, because we can garner large support for our beliefs in hypotheses based on evidence, it seems plausible that a machine could as well. Furthermore, a machine can be justified in their degrees of belief, insofar as their programmers have given them sufficient and unbiased data, and that their algorithms used to examine that data are well formed. Consider a computer learning how to identify a dog from a picture. Just as human children learn what a dog is by observing one and being told by another human that that's a dog, machines can examine a picture via artificial neural networks and estimate whether or not that picture has a dog in it. As the programmers act as the judge, telling the machine whether they were right or wrong in their estimation, the machine adjusts its degree of belief and understanding of what a dog looks like. Now, if we take a Bayesian twist on the classical definition of knowledge, we could say that in order to have knowledge in something that thing must be true, you must believe it to a large degree, and you must be justified in your degree of belief. Thus, it follows that a machine can acquire knowledge in this Bayesian sense of the word.

In this paper I've asked the question whether or not machines can acquire knowledge, and if that knowledge is the same as human knowledge. To answer this I first considered Hume's Problem of Induction and then examined the Bayesian answer to it. Taking the Bayesian position I concluded that a machine can indeed acquire knowledge in a certain sense of the word, and that that knowledge is similar at some level to human knowledge. As research in AI continues questions regarding the nature of the knowledge machines can acquire remain important to consider. It is essential that we continue to explore this union of epistemology and computer science in order to better understand the technology we are creating.