A Reply to Hellerstein's Book Review of
Machine Learning: A Theoretical Approach

B.K. NATARAJAN
Hewlett Packard Laboratories, 1501 Page Mill Road, Palo Alto, CA 94304

I want to thank Lisa Hellerstein for her kind efforts, and Alberto Segre for commissioning such.

The challenge in writing this book lay in presenting technically difficult material in a form that was accessible to the broader audience, without compromising its integrity. A compounding factor was that this was to be the first text in the area, without prior leverage in establishing notational conventions or conceptual organization. Such a challenge demanded numerous editorial and pedagogic decisions, and I must confess that many of Hellerstein's points bear directly on these decisions. While it is neither possible nor constructive for me to respond to her review exhaustively, I would urge anyone undertaking a similar work to pay her close attention. I will restrict my response to a few specific points.

When I say that the PAC model appears to be a good model of the natural learning process, I do not imply that it models the learning mechanisms of the living psyche. I mean that it is a good model of the input–output behavior of the natural learning process—good because it is worth studying in an aesthetic sense, and more importantly, good because it promises to be a step towards explaining human experience. As pointed out by Valiant (1984), both of the above qualities are desirable in a computational model. My statement is in the same spirit as the statement that the Turing machine is a good model of the natural notion of a computational algorithm.

I am pleased that Hellerstein sees my book as a uniform presentation of theoretical results, offering scant press to philosophical issues. That was my intention. It is not clear to me that the PAC model has been studied to the extent necessary to support a philosophical dissection, and even if such were the case, that would be best left to someone more didactically mature than myself.

Hellerstein correctly points out that chapter 7 is brief on motivating the study of neural nets, and brief on the implications of the results presented. The brevity is intentional and can be traced to two factors. Firstly, the topic of neural nets enjoys considerable attention in the literature, and there are numerous sources of introductory material. It did not seem fruitful for me to attempt supplanting them. Secondly, there is vigorous debate in the scientific community as to the nature and contribution of neural networks. The results presented in chapter 7 stand alone on their technical merit. Rather than obfuscate them with an attempt at joining the debate in the broader context, I chose to leave it to the reader or the instructor to interpret the results as they see fit.

Hellerstein asks whether the discussion of the Vapnik–Chervonenkis dimension should have been postponed to chapter 4, since countable concept classes can be handled by simple

counting arguments. This is a good pedagogic question. Earlier versions of the manuscript did exactly this, before it was evident that a significant fraction of the audience felt cheated by counting arguments—they were eager to circumnavigate the heavy machinery of chapter 4 and yet gain some facility with the Vapnik–Chervonenkis dimension.

In all, the aim of my book is to provide a core collection of theoretical results in a form that is accessible to the broader audience. The aim is not to be comprehensive, for this is all but futile in a rapidly growing discipline, and not to be philosophical, since that seems premature. On the one hand, Hellerstein's review is congratulatory with regard to my stated aims, and I am grateful for her verdict. On the other hand, she identifies areas that demand additional coverage or improvement, and I sincerely hope that these will be addressed by later books.

References

Valiant, L.G. (1984). A theory of the learnable. *Communications of the ACM*, 27, 1134–1142.