

This is a section of [doi:10.7551/mitpress/14723.001.0001](https://doi.org/10.7551/mitpress/14723.001.0001)

Gradient Expectations

Structure, Origins, and Synthesis of Predictive Neural Networks

By: Keith L. Downing

Citation:

Gradient Expectations: Structure, Origins, and Synthesis of Predictive Neural Networks

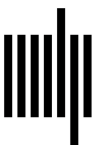
By: Keith L. Downing

DOI: 10.7551/mitpress/14723.001.0001

ISBN (electronic): 9780262374675

Publisher: The MIT Press

Published: 2023



The MIT Press

Notes

1. Introduction

1. In baseball, a strikeout occurs when a batter misses on three attempts at hitting a pitched ball, while a round-tripper is a slang expression for a homerun, in which the batter hits the ball (typically very far) and is able to run around all bases and back to the starting base (home) to score one or more points, depending on the number of teammates who were on the bases at the time.

2. I experience this temporal mismatch when standing in an amusement park's baseball batting cage in which the velocity of incoming balls has been set very high. The approaching ball is visible, albeit blurry, but I cannot even begin to react before it strikes the backstop. My body freezes, totally dumbfounded by it all.

2. Conceptual Foundations of Prediction

1. The more common term is *allocentric*, but its standard connotation involves the perspective of other people, while its philosophical and scientific meaning is *from the perspective of anything other than the individual*.

2. It is hard to read a DL article without encountering multiple occurrences of the word *gradient*.

3. The direction of people's mental number line tends to mirror the direction in which they read. Hence, most native English speakers view larger numbers to the right, while Arabic speakers envision them on the left (Dehaene 1997).

4. Notice that this is actually three nested averages: (1) Each student's GPA is the average grade over all of their courses. (2) These student' GPAs are averaged to create the mathematics department's average for its incoming student group for a particular year. (3) The year-group averages are averaged to make a prediction. The first two averages are normally unbiased, while the third may use unequal weights to improve predictive accuracy.

5. It seems reasonable to assume that $k_g + k_a = 1$, but forgoing this constraint helps highlight the similarities between prediction and control.

6. Point guard and forward are roles performed by basketball players.

3. Biological Foundations of Prediction

1. In general, sophisticated motor activity demands widespread inhibition. In fact, local actions often require more restraint and thus constitute more complex activities, as mentioned earlier about moving one versus all fingers or toes.

4. Neural Energy Networks

1. Hopfield's original neurons used 0 and 1 but essentially scaled them to -1 or 1 at the start of most calculations.

2. In the original nomenclature, an RBM consists of one pair of layers, while the full network comprises a stack of RBMs, with one variant known as a deep belief network (DBN). However, other literature uses RBM more generically, as either one or many paired layers. This book uses the generic connotation of RBM.

3. The term $p_g(s|d)$ is an odd conditional probability, since the generative weights produce data (d) from internal states (s), not s from d. Simple use of Bayes rule expresses this in terms of three probabilities that make more intuitive sense: $p_g(s|d) = \frac{p_g(d|s)p_g(s)}{p_g(d)}$. Here, $p_g(s)$ are just the a priori internal-state probabilities.

4. This would make a nice motto for an alternative-living commune.

5. Predictive Coding

1. Of course, the value of $F'(s_k)$ depends on the activation function, F . When F is linear, $F'(S_k) = 1$, while the logistic activation function (aka sigmoid) has $F'(S_k) = F(S_k)(1 - F(S_k))$, and the hyperbolic tangent has $F'(S_k) = 1 - F(S_k)^2$.

2. The negative signs in these rules assume an objective function of positive magnitude (e.g., free energy) that predictive coding strives to minimize. Alternatively, Whittington and Bogacz (2017) maximize a negative energy function and thus have no negative signs in their learning rules.

3. The brain accounts for roughly 2 percent of the body's mass but uses about 20 percent of its energy (Sterling and Laughlin 2015).

6. Emergence of Predictive Networks

1. The expression *ontogeny recapitulates phylogeny* implies that key evolutionary transitions in the ancestral lineage of a species are replayed (recapitulated) in its embryology.

2. Some of my neuroscientist colleagues do not even recognize the concept of a neuromodulator; they view all of these chemicals as neurotransmitters that can be released into the synaptic cleft or into wider expanses of the brain.

7. Evolving Artificial Predictive Networks

1. This is a reference to the often-cited phrase of uncertain origins, *If all you have is a hammer, everything looks like a nail*.

2. In the article, they actually maximize the negative of FEP, also known as the *evidence lower bound* (ELBO).

3. Much of D'Arcy was inspired by the background reading for this book, and although fully implemented, the system is still in the early phases of testing.

8. Conclusion

1. Prior to the 2010s, this was an accusation that I often received, on behalf of AI, from engineers in other departments. Now, those same departments are creating their own AI professorships.

© 2023 Keith L. Downing

This work is subject to a Creative Commons CC-BY-NC-ND license. Subject to such license, all rights are reserved.



The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Times New Roman by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Downing, Keith L., author.

Title: Gradient expectations : structure, origins, and synthesis of predictive neural networks / Keith L. Downing.

Description: [Cambridge, Massachusetts] : The MIT Press, [2023] | Includes bibliographical references and index.

Identifiers: LCCN 2022037237 (print) | LCCN 2022037238 (ebook) |

ISBN 9780262545617 (paperback) | ISBN 9780262374682 (epub) |

ISBN 9780262374675 (pdf)

Subjects: LCSH: Deep learning (Machine learning) | Neural networks (Computer science) | Conjugate gradient methods.

Classification: LCC Q325.73 .D88 2023 (print) | LCC Q325.73 (ebook) |

DDC 006.3/2—dc23/eng20230302

LC record available at <https://lcn.loc.gov/2022037237>

LC ebook record available at <https://lcn.loc.gov/2022037238>