

---

Sandro Skansi

# Introduction to Deep Learning

From Logical Calculus to Artificial Intelligence

---

# Contents

|          |  |    |
|----------|--|----|
| <b>1</b> | <b>From Logic to Cognitive Science</b> . . . . .                       | 1  |
| 1.1      | The Beginnings of Artificial Neural Networks . . . . .                 | 1  |
| 1.2      | The XOR Problem . . . . .  | 5  |
| 1.3      | From Cognitive Science to Deep Learning . . . . .                      | 8  |
| 1.4      | Neural Networks in the General AI Landscape . . . . .                  | 11 |
| 1.5      | Philosophical and Cognitive Aspects . . . . .                          | 12 |
|          | References . . . . .   | 15 |
| <b>2</b> | <b>Mathematical and Computational Prerequisites</b> . . . . .          | 17 |
| 2.1      | Derivations and Function Minimization . . . . .                        | 17 |
| 2.2      | Vectors, Matrices and Linear Programming . . . . .                     | 25 |
| 2.3      | Probability Distributions . . . . .                                    | 32 |
| 2.4      | Logic and Turing Machines . . . . .                                    | 39 |
| 2.5      | Writing Python Code . . . . .  | 41 |
| 2.6      | A Brief Overview of Python Programming . . . . .                       | 43 |
|          | References . . . . .   | 49 |
| <b>3</b> | <b>Machine Learning Basics</b> . . . . .                               | 51 |
| 3.1      | Elementary Classification Problem . . . . .                            | 51 |
| 3.2      | Evaluating Classification Results . . . . .                            | 57 |
| 3.3      | A Simple Classifier: Naive Bayes . . . . .                             | 59 |
| 3.4      | A Simple Neural Network: Logistic Regression . . . . .                 | 61 |
| 3.5      | Introducing the MNIST Dataset . . . . .                                | 68 |
| 3.6      | Learning Without Labels: K-Means . . . . .                             | 70 |
| 3.7      | Learning Different Representations: PCA . . . . .                      | 72 |
| 3.8      | Learning Language: The Bag of Words Representation . . . . .           | 75 |
|          | References . . . . .   | 77 |
| <b>4</b> | <b>Feedforward Neural Networks</b> . . . . .                           | 79 |
| 4.1      | Basic Concepts and Terminology for Neural Networks . . . . .           | 79 |
| 4.2      | Representing Network Components with Vectors<br>and Matrices . . . . . | 82 |
| 4.3      | The Perceptron Rule . . . . .  | 84 |
| 4.4      | The Delta Rule . . . . .   | 87 |

|          |   |            |
|----------|---|------------|
| 4.5      | From the Logistic Neuron to Backpropagation . . . . .                               | 89         |
| 4.6      | Backpropagation . . . . .   | 93         |
| 4.7      | A Complete Feedforward Neural Network . . . . .                                     | 102        |
|          | References. . . . .   | 105        |
| <b>5</b> | <b>Modifications and Extensions to a Feed-Forward Neural Network . . . . .</b>      | <b>107</b> |
| 5.1      | The Idea of Regularization . . . . .  | 107        |
| 5.2      | $L_1$ and $L_2$ Regularization . . . . .  | 109        |
| 5.3      | Learning Rate, Momentum and Dropout. . . . .  | 111        |
| 5.4      | Stochastic Gradient Descent and Online Learning . . . . .                           | 116        |
| 5.5      | Problems for Multiple Hidden Layers: Vanishing<br>and Exploding Gradients . . . . . | 118        |
|          | References. . . . .   | 119        |
| <b>6</b> | <b>Convolutional Neural Networks . . . . .</b>                                      | <b>121</b> |
| 6.1      | A Third Visit to Logistic Regression . . . . .                                      | 121        |
| 6.2      | Feature Maps and Pooling . . . . .  | 125        |
| 6.3      | A Complete Convolutional Network. . . . .   | 127        |
| 6.4      | Using a Convolutional Network to Classify Text . . . . .                            | 130        |
|          | References. . . . .   | 132        |
| <b>7</b> | <b>Recurrent Neural Networks . . . . .</b>  | <b>135</b> |
| 7.1      | Sequences of Unequal Length . . . . .   | 135        |
| 7.2      | The Three Settings of Learning with Recurrent Neural<br>Networks . . . . .          | 136        |
| 7.3      | Adding Feedback Loops and Unfolding a Neural Network . . . . .                      | 139        |
| 7.4      | Elman Networks . . . . .  | 140        |
| 7.5      | Long Short-Term Memory . . . . .  | 142        |
| 7.6      | Using a Recurrent Neural Network for Predicting<br>Following Words . . . . .        | 145        |
|          | References. . . . .   | 152        |
| <b>8</b> | <b>Autoencoders . . . . .</b>   | <b>153</b> |
| 8.1      | Learning Representations . . . . .  | 153        |
| 8.2      | Different Autoencoder Architectures. . . . .  | 156        |
| 8.3      | Stacking Autoencoders . . . . .   | 158        |
| 8.4      | Recreating the Cat Paper . . . . .  | 161        |
|          | References. . . . .   | 163        |
| <b>9</b> | <b>Neural Language Models . . . . .</b>   | <b>165</b> |
| 9.1      | Word Embeddings and Word Analogies. . . . .   | 165        |
| 9.2      | CBOW and Word2vec . . . . .   | 166        |
| 9.3      | Word2vec in Code . . . . .  | 168        |

---

|           |  |            |
|-----------|--|------------|
| 9.4       | Walking Through the Word-Space: An Idea That<br>Has Eluded Symbolic AI . . . . . | 171        |
|           | References. . . . .  | 173        |
| <b>10</b> | <b>An Overview of Different Neural Network Architectures. . . . .</b>            | <b>175</b> |
| 10.1      | Energy-Based Models. . . . .   | 175        |
| 10.2      | Memory-Based Models. . . . .   | 178        |
| 10.3      | The Kernel of General Connectionist Intelligence:<br>The bAbI Dataset . . . . .  | 181        |
|           | References. . . . .  | 182        |
| <b>11</b> | <b>Conclusion . . . . .</b>  | <b>185</b> |
| 11.1      | An Incomplete Overview of Open Research Questions . . . . .                      | 185        |
| 11.2      | The Spirit of Connectionism and Philosophical Ties . . . . .                     | 186        |
|           | Reference . . . . .  | 187        |
|           | <b>Index . . . . .</b>   | <b>189</b> |