

## Personalized Machine Learning

Every day we interact with machine learning systems offering individualized predictions for our entertainment, social connections, purchases, or health. These involve several modalities of data, from sequences of clicks to text, images, and social interactions. This book introduces common principles and methods that underpin the design of personalized predictive models for a variety of settings and modalities.

The book begins by revising ‘traditional’ machine learning models, focusing on how to adapt them to settings involving user data; then presents techniques based on advanced principles such as matrix factorization, deep learning, and generative modeling; and concludes with a detailed study of the consequences and risks of deploying personalized predictive systems.

A series of case studies in domains ranging from e-commerce to health plus hands-on projects and code examples will give readers understanding and experience with large-scale real-world datasets and the ability to design models and systems for a wide range of applications.

JULIAN MCAULEY has been a Professor at the University of California San Diego since 2014. Personalized Machine Learning is the main research area of his lab, with applications ranging from personalized recommendation to dialog, health care, and fashion design. He regularly collaborates with industry on these topics, including Amazon, Facebook, Microsoft, Salesforce, and Etsy. His work has been selected for several awards including an NSF CAREER award, and faculty awards from Amazon, Salesforce, Facebook, and Qualcomm, among others.

Cambridge University Press  
978-1-316-51890-8 — Personalized Machine Learning  
Julian McAuley  
Frontmatter  
[More Information](#)

---

# Personalized Machine Learning

JULIAN MCAULEY  
*University of California San Diego*



Cambridge University Press  
978-1-316-51890-8 — Personalized Machine Learning  
Julian McAuley  
Frontmatter  
[More Information](#)

---

**CAMBRIDGE**  
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New  
Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 23846

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of  
education, learning, and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781316518908](http://www.cambridge.org/9781316518908)

DOI: 10.1017/9781009003971

© Julian McAuley 2022

This publication is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without the written  
permission of Cambridge University Press.

First published 2022

*A catalogue record for this publication is available from the British Library.*

ISBN 978-1-316-51890-8 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of  
URLs for external or third-party internet websites referred to in this publication  
and does not guarantee that any content on such websites is, or will remain,  
accurate or appropriate.

Contents

|          | <i>Notation</i>                                    | <i>page ix</i> |
|----------|--|----------------|
| <b>1</b> | <b>Introduction</b>                                | 1              |
|          | 1.1 Purpose of This Book                           | 2              |
|          | 1.2 For Learners: What Is Covered, and What Is Not | 3              |
|          | 1.3 For Instructors: Course and Content Outline    | 5              |
|          | 1.4 Online Resources                               | 7              |
|          | 1.5 About the Author                               | 8              |
|          | 1.6 Personalization in Everyday Life               | 9              |
|          | 1.7 Techniques for Personalization                 | 12             |
|          | 1.8 The Ethics and Consequences of Personalization | 14             |
|          | <b>PART ONE MACHINE LEARNING PRIMER</b>            | 17             |
| <b>2</b> | <b>Regression and Feature Engineering</b>          | 19             |
|          | 2.1 Linear Regression                              | 21             |
|          | 2.2 Evaluating Regression Models                   | 26             |
|          | 2.3 Feature Engineering                            | 32             |
|          | 2.4 Interpreting the Parameters of Linear Models   | 40             |
|          | 2.5 Fitting Models with Gradient Descent           | 42             |
|          | 2.6 Nonlinear Regression                           | 43             |
|          | Exercises  | 46             |
|          | Project 1: Taxicab Tip Prediction (Part 1)         | 47             |
| <b>3</b> | <b>Classification and the Learning Pipeline</b>    | 49             |
|          | 3.1 Logistic Regression                            | 50             |
|          | 3.2 Other Classification Techniques                | 52             |
|          | 3.3 Evaluating Classification Models               | 54             |
|          | 3.4 The Learning Pipeline                          | 62             |

|   |  |     |
|---|--|-----|
| 3.5   | Implementing the Learning Pipeline                                     | 72  |
|   | Exercises  | 76  |
|   | Project 2: Taxicab Tip Prediction (Part 2)                             | 77  |
| <b>PART TWO FUNDAMENTALS OF PERSONALIZED MACHINE LEARNING</b> |  | 79  |
| <b>4</b>  | <b>Introduction to Recommender Systems</b>                             | 81  |
| 4.1   | Basic Setup and Problem Definition                                     | 82  |
| 4.2   | Representations for Interaction Data                                   | 84  |
| 4.3   | Memory-Based Approaches to Recommendation                              | 86  |
| 4.4   | Random Walk Methods  | 97  |
| 4.5   | Case Study: Amazon.com Recommendations                                 | 100 |
|   | Exercises  | 100 |
|   | Project 3: A Recommender System for Books (Part 1)                     | 102 |
| <b>5</b>  | <b>Model-Based Approaches to Recommendation</b>                        | 104 |
| 5.1   | Matrix Factorization   | 106 |
| 5.2   | Implicit Feedback and Ranking Models                                   | 112 |
| 5.3   | ‘User-free’ Model-Based Approaches                                     | 117 |
| 5.4   | Evaluating Recommender Systems   | 121 |
| 5.5   | Deep Learning for Recommendation                                       | 125 |
| 5.6   | Retrieval  | 131 |
| 5.7   | Online Updates   | 133 |
| 5.8   | Recommender Systems in Python with <i>Surprise</i> and <i>Implicit</i> | 134 |
| 5.9   | Beyond a ‘Black-Box’ View of Recommendation                            | 138 |
| 5.10  | History and Emerging Directions  | 139 |
|   | Exercises  | 141 |
|   | Project 4: A Recommender System for Books (Part 2)                     | 142 |
| <b>6</b>  | <b>Content and Structure in Recommender Systems</b>                    | 144 |
| 6.1   | The Factorization Machine  | 145 |
| 6.2   | Cold-Start Recommendation  | 149 |
| 6.3   | Multisided Recommendation  | 152 |
| 6.4   | Group- and Socially Aware Recommendation                               | 155 |
| 6.5   | Price Dynamics in Recommender Systems                                  | 162 |
| 6.6   | Other Contextual Features in Recommendation                            | 167 |
| 6.7   | Online Advertising   | 170 |
|   | Exercises  | 174 |
|   | Project 5: Cold-Start Recommendation on <i>Amazon</i>                  | 175 |

|           | <i>Contents</i>  | vii |
|-----------|--|-----|
| <b>7</b>  | <b>Temporal and Sequential Models</b>  | 177 |
| 7.1       | Introduction to Regression with Time Series                                  | 178 |
| 7.2       | Temporal Dynamics in Recommender Systems                                     | 180 |
| 7.3       | Other Approaches to Temporal Dynamics  | 188 |
| 7.4       | Personalized Markov Chains   | 191 |
| 7.5       | Case Studies: Markov-Chain Models<br>for Recommendation                      | 193 |
| 7.6       | Recurrent Networks   | 201 |
| 7.7       | Neural Network-Based Sequential Recommenders                                 | 204 |
| 7.8       | Case Study: Personalized Heart-Rate Modeling                                 | 210 |
| 7.9       | History of Personalized Temporal Models                                      | 211 |
|           | Exercises  | 212 |
|           | Project 6: Temporal and Sequential Dynamics in Business<br>Reviews           | 213 |
|           | <b>PART THREE EMERGING DIRECTIONS IN PERSON-<br/>ALIZED MACHINE LEARNING</b> | 217 |
| <b>8</b>  | <b>Personalized Models of Text</b>   | 219 |
| 8.1       | Basics of Text Modeling: The Bag-of-Words Model                              | 220 |
| 8.2       | Distributed Word and Item Representations                                    | 230 |
| 8.3       | Personalized Sentiment and Recommendation                                    | 233 |
| 8.4       | Personalized Text Generation   | 237 |
| 8.5       | Case Study: <i>Google's Smart Reply</i>                                      | 247 |
|           | Exercises  | 249 |
|           | Project 7: Personalized Document Retrieval                                   | 250 |
| <b>9</b>  | <b>Personalized Models of Visual Data</b>                                    | 252 |
| 9.1       | Personalized Image Search and Retrieval                                      | 253 |
| 9.2       | Visually Aware Recommendation and Personalized<br>Ranking                    | 254 |
| 9.3       | Case Studies: Visual and Fashion Compatibility                               | 257 |
| 9.4       | Personalized Generative Models of Images                                     | 267 |
|           | Exercises  | 269 |
|           | Project 8: Generating Compatible Outfits                                     | 271 |
| <b>10</b> | <b>The Consequences of Personalized Machine Learning</b>                     | 273 |
| 10.1      | Measuring Diversity  | 275 |
| 10.2      | Filter Bubbles, Diversity, and Extremification                               | 277 |
| 10.3      | Diversification Techniques   | 279 |
| 10.4      | Implementing a Diverse Recommender   | 284 |

|      |  |     |
|------|--|-----|
| viii | <i>Contents</i>                                |     |
| 10.5 | Case Studies on Recommendation and Consumption |     |
|      | Diversity                                      | 286 |
| 10.6 | Other Metrics Beyond Accuracy                  | 291 |
| 10.7 | Fairness                                       | 295 |
| 10.8 | Case Studies on Gender Bias in Recommendation  | 300 |
|      | Exercises                                      | 303 |
|      | Project 9: Diverse and Fair Recommendations    | 305 |
|      | <i>References</i>                              | 306 |
|      | <i>Index</i>                                   | 322 |



# Notation

## Common Mathematical Symbols

### Machine Learning

|                     |  |
|---------------------|--|
| $y$                 | vector of labels   |
| $X$                 | matrix of features   |
| $x_i$               | feature vector for the $i$ th sample   |
| $f(x_i)$            | model prediction for the $i$ th sample   |
| $r_i$               | residual (error) associated with the $i$ th prediction, $r_i = (y_i - f(x_i))$ |
| $\theta$            | vector of model parameters   |
| $\sigma$            | sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$                              |
| $\ x\ _p$           | $p$ -norm, $\ x\ _p = (\sum_i  x_i ^p)^{1/p}$                                  |
| $\ell_1; \ell_2$    | regularizers $\ \theta\ _1$ and $\ \theta\ _2$                                 |
| $\lambda$           | regularization hyperparameter  |
| $\mathcal{L}; \ell$ | likelihood and log-likelihood  |

### Users and Items

|            |   |
|------------|---|
| $u \in U$  | user $u$ in user set $U$  |
| $i \in I$  | item $i$ in item set $I$  |
| $I_u$      | set of items rated (or interacted with) by user $u$                                       |
| $U_i$      | set of users who have rated (or interacted with) item $i$                                 |
| $ U ;  I $ | number of users and number of items   |
| $R_{u,i}$  | measurement (e.g., a rating) associated with an interaction between user $u$ and item $i$ |
| $x_{u,i}$  | model estimate of the compatibility between user $u$ and item $i$                         |

### Recommender Systems

|           |                                    |
|-----------|------------------------------------|
| $\beta_u$ | bias term associated with user $u$ |
| $\beta_i$ | bias term associated with item $i$ |

| x                    | <i>Notation</i>                                      |
|----------------------|--|
| $\gamma_u$           | vector of parameters describing a single user $u$    |
| $\gamma_i$           | vector of parameters describing a single item $i$    |
| $\gamma_U; \gamma_I$ | parameters for all users $U$ or all items $I$        |
| $K$                  | feature dimensionality (or number of latent factors) |

Common Abbreviations

|        |   |
|--------|---|
| AUC    | area under the ROC curve (eq. (5.26))                     |
| BER    | balanced error rate (eq. (3.20))                          |
| BPR    | Bayesian personalized ranking (sec. 5.2.2)                |
| CNN    | convolutional neural network (sec. 5.5.4)                 |
| FVU    | fraction of variance unexplained (eq. (2.32))             |
| FN/FNR | false negatives/false negative rate (sec. 3.3.1)          |
| FP/FPR | false positives/false positive rate (sec. 3.3.1)          |
| GAN    | generative adversarial network (sec. 9.4)                 |
| LSTM   | long short-term memory model (sec. 7.6)                   |
| MAE    | mean absolute error (eq. (2.17))                          |
| MLE    | maximum likelihood estimation (sec. 2.2.3)                |
| MLP    | multilayer perceptron (sec. 5.5.2)                        |
| MMR    | maximal marginal relevance (sec. 10.3.1)                  |
| MRR    | mean reciprocal rank (sec. 5.4.2)                         |
| MSE    | mean squared error (sec. 2.2.1)                           |
| NDCG   | normalized discounted cumulative gain (sec. 5.4.3)        |
| RNN    | recurrent neural network (sec. 7.6)                       |
| ROC    | receiver-operating characteristic (sec. 3.3.3)            |
| SVM    | support vector machine (sec. 3.2)                         |
| TF-IDF | term frequency and inverse document frequency (eq. (8.8)) |
| TN/TNR | true negatives/true negative rate (sec. 3.3.1)            |
| TP/TPR | true positives/true positive rate (sec. 3.3.1)            |