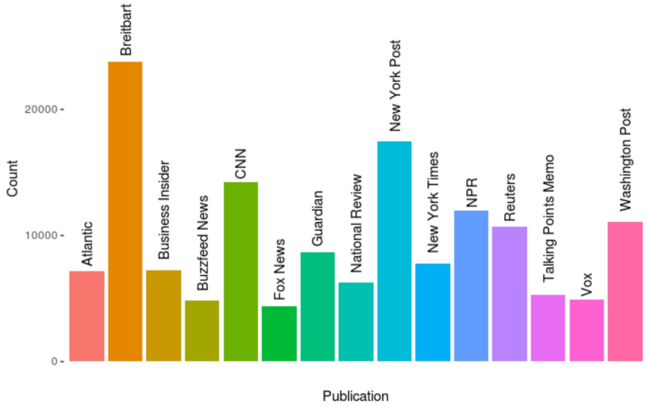
Recommendation as Search/Information Retrieval

Example: Recommending News Articles

- Dataset: https://www.kaggle.com/snapcrack/all-the-news/home
- Fields: id, title, publication name, author, date, year, month, url, content
- We'll use first 1,000 articles from articles1.csv (which contains 50,000 articles)

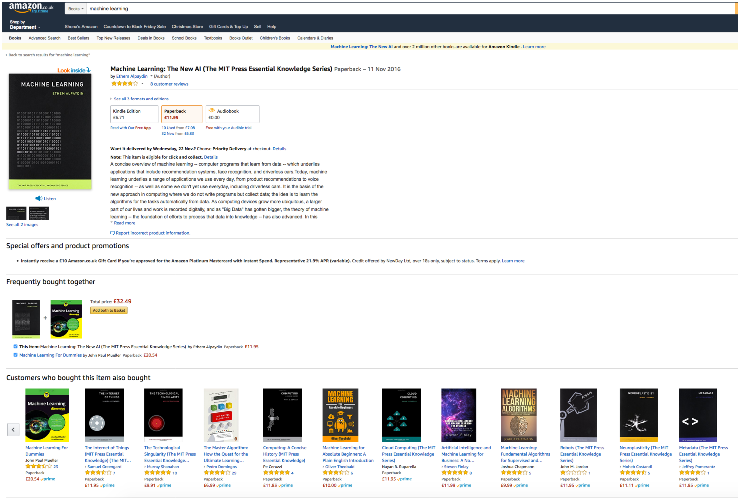


Example: Recommending News Articles

- Remove stop words, use steming
- Bag of words model
- Use 500 most frequent terms (to speed things up)
- Given an article, find *k* nearest neighbours and Euclidean distance
- e.g. article "House Republicans Fret About Winning Their Health Care Suit':
    - House Clears Path for Repeal of Health Law 0.82
    - Trump Follows Obamas Lead in Flexing Executive Muscle 0.83
    - Senate Republicans Open Fight Over Obama Health Law 0.85
    - G.O.P. Campaign to Repeal Obamacare Stalls on the Details 0.85
    - Turmoil Overshadows First Day of Republican-Controlled Congress 0.86

Example: Recommending Shopping

- Feature vector of length N, the number of different items stocked
- For "shopping basket" set feature vector entry $=1$ if item in basket
- To make recommendations use kNN, then rank items by popularity


amazon.co.uk Try Prime | Books | machine learning

Shop by Department | Shona's Amazon | Countdown to Black Friday Sale | Christmas Store | Gift Cards & Top Up | Sell | Help

Books | Advanced Search | Best Sellers | Top New Releases | Deals in Books | School Books | Textbooks | Books Outlet | Children's Books | Calendars & Diaries

Machine Learning: The New AI and over 2 million other books are available for **Amazon Kindle** . Learn more

‹ Back to search results for "machine learning"

Look inside

Machine Learning: The New AI (The MIT Press Essential Knowledge Series) Paperback – 11 Nov 2016

by Ethem Alpaydin (Author)

8 customer reviews

See all 3 formats and editions

Kindle Edition	Paperback	Audiobook
£6.71	£11.95	£0.00
Read with Our Free App	10 Used from £7.08 32 New from £6.83	Free with your Audible trial

Want it delivered by Wednesday, 22 Nov.? Choose **Priority Delivery** at checkout. Details

Note: This item is eligible for **click and collect.** Details

A concise overview of machine learning -- computer programs that learn from data -- which underlies applications that include recommendation systems, face recognition, and driverless cars.Today, machine learning underlies a range of applications we use every day, from product recommendations to voice recognition -- as well as some we don't yet use everyday, including driverless cars. It is the basis of the new approach in computing where we do not write programs but collect data; the idea is to learn the algorithms for the tasks automatically from data. As computing devices grow more ubiquitous, a larger part of our lives and work is recorded digitally, and as "Big Data" has gotten bigger, the theory of machine learning -- the foundation of efforts to process that data into knowledge -- has also advanced. In this

Read more

Listen

See all 2 images

Report incorrect product information.

Special offers and product promotions

- **Instantly receive a £10 Amazon.co.uk Gift Card if you're approved for the Amazon Platinum Mastercard with Instant Spend. Representative 21.9% APR (variable).** Credit offered by NewDay Ltd, over 18s only, subject to status. Terms apply. Learn more

Frequently bought together

Total price: £32.49

Add both to Basket

☑ **This item:** Machine Learning: The New AI (The MIT Press Essential Knowledge Series) by Ethem Alpaydin Paperback £11.95

☑ Machine Learning For Dummies by John Paul Mueller Paperback £20.54

Customers who bought this item also bought

Title	Author	Reviews	Format	Price
Machine Learning For Dummies	John Paul Mueller	23	Paperback	£20.54 prime
The Internet of Things (MIT Press Essential Knowledge) (The MIT…	Samuel Greengard	7	Paperback	£11.95 prime
The Technological Singularity (The MIT Press Essential Knowledge…	Murray Shanahan	10	Paperback	£9.91 prime
The Master Algorithm: How the Quest for the Ultimate Learning…	Pedro Domingos	29	Paperback	£6.99 prime
Computing: A Concise History (MIT Press Essential Knowledge)	Pe Ceruzzi	4	Paperback	£11.83 prime
Machine Learning for Absolute Beginners: A Plain English Introduction	Oliver Theobald	6	Paperback	£10.00 prime
Cloud Computing (The MIT Press Essential Knowledge Series)	Nayan B. Ruparelia		Paperback	£11.95 prime
Artificial Intelligence and Machine Learning for Business: A No…	Steven Finlay	8	Paperback	£11.99 prime
Machine Learning: Fundamental Algorithms for Supervised and…	Joshua Chapmann	5	Paperback	£9.99 prime
Robots (The MIT Press Essential Knowledge Series)	John M. Jordan	1	Paperback	£11.95 prime
Neuroplasticity (The MIT Press Essential Knowledge Series)	Moheb Costandi	5	Paperback	£11.11 prime
Metadata (The MIT Press Essential Knowledge Series)	Jeffrey Pomerantz	1	Paperback	£11.95 prime

Collaborative Filtering Recommenders

- Search-based recommendations cheap and effective but have no personalisation, no surprises/exploration
- Main idea:
    - Store history (articles read, items bought, songs listened to etc) for each user, plus ratings if available
    - To make recommendations for a user find similar users and use their histories to form recommendations
- E.g. use kNN again to find nearest users, then recommend most popular items from the set of nearest users (weighted by user ratings of items, if available)
- Two major problems:
    - Sparse Data: User history data is usually v sparse (most users only rate a few items e.g. 10 items out of a possible 100k items) $\rightarrow$ hard to reliably find neighbouring users unless have a lot of data
    - Cold Start: New users have no history, new items are not in any users history

Collaborative Filtering As Matrix Completion

Example: users rate books they have read from 0-5.

Book	Alice(1)	Bob(2)	Carol(3)	Dave(4)
Machine Learning for Dummies(1)	5	4	0	0
Hands-On Machine Learning(2)	?	5	?	0
Deep Learning(3)	5	?	?	?
A Kitten Called Holly(4)	0	0	5	?
Kittens 2018 Calendar(5)	0	0	5	4

Notation:

- n number of users, $n = 4$
- m number of items, $m = 5$
- d number of features
- R_{uv} rating given by user u to item v, $R_{11} = 5$
- $\delta_{uv} = 1$ if item v rated by user u, 0 otherwise, $\delta_{11} = 1$, $\delta_{12} = 0$

Matrix Completion

Book	Alice(1)	Bob(2)	Carol(3)	Dave(4)
Machine Learning for Dummies(1)	5	4	0	0
Hands-On Machine Learning(2)	?	5	?	0
Deep Learning(3)	5	?	?	?
A Kitten Called Holly(4)	0	0	5	?
Kittens 2018 Calendar(5)	0	0	5	4

- Associate a feature vector $x^{(v)}$ with v'th book, e.g. $x^{(1)} = [1, 0]^T$, $x^{(4)} = [0, 1]^T$ (number of features $d = 2$)
- For each user u learn parameter vector $\theta^{(u)}$, e.g. $\theta^{(1)} = [5, 0]^T$, $\theta^{(3)} = [0, 5]^T$
- Predicted rating by user u of item v is $(\theta^{(u)})^T x$, e.g. rating by user 1 of item 1 its $[5, 0] \times [1, 0]^T = 5$

Matrix Completion

- We are given a feature vector $x^{(v)}$ for v'th item/book
- Training data: a set of ratings $\{R_{uv}\}$ by users of a subset of the items (each user might only rate a few items)
- Hypothesis: predicted rating by user u of item v is: $h_{\theta^{(u)}}(x^{(v)}) = (\theta^{(u)})^T x^{(v)}$
- Parameters: $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n)}$
- Cost function:

  $J(\theta^{(1)}, \dots, \theta^{(n)}) = \sum_{u=1}^{n} \sum_{v=1}^{m} \delta_{uv}(R_{uv} - (\theta^{(u)})^T x^{(v)})^2 + \lambda \sum_{u=1}^{n} (\theta^{(u)})^T \theta^{(u)}$
- Select $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n)}$ to minimise this cost function. This requires solving a least squares problem: use gradient descent or closed-form solution.

Matrix Completion

Book	Alice(1)	Bob(2)	Carol(3)	Dave(4)	$x^{(1)}$=[ML	kittens]
Machine Learning for Dummies(1)	5	4	0	0	1	0
Hands-On Machine Learning(2)	?	5	?	0	?	?
Deep Learning(3)	5	?	?	?	?	?
A Kitten Called Holly(4)	0	0	5	?	?	?
Kittens 2018 Calendar(5)	0	0	5	4	0	1

- Associate a feature vector $x^{(v)}$ with v'th book. But what if we don't know $x^{(v)}$?
- Suppose we know $\theta^{(1)} = [5, 0]^T$, $\theta^{(3)} = [0, 5]^T$, then

$$[5,0]^T x^{(1)} = 5,\ [5,0]^T x^{(3)} = 5,\ [5,0]^T x^{(4)} = 0$$
$$[0,5]^T x^{(1)} = 0,\ [0,5]^T x^{(4)} = 5$$

which is satisfied by $x^{(1)} = [1, 0]$, $x^{(3)} = [1, 0]$, $x^{(4)} = [0, 1]$

Matrix Completion

- Given $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(n)}$, select $x^{(1)}, x^{(2)}, \ldots, x^{(m)}$ to minimise

$$\sum_{u=1}^{n}\sum_{v=1}^{m}\delta_{uv}(R_{uv} - (\theta^{(u)})^T x^{(v)})^2 + \lambda\sum_{v=1}^{m}(x^{(v)})^T x^{(v)}$$

- Define

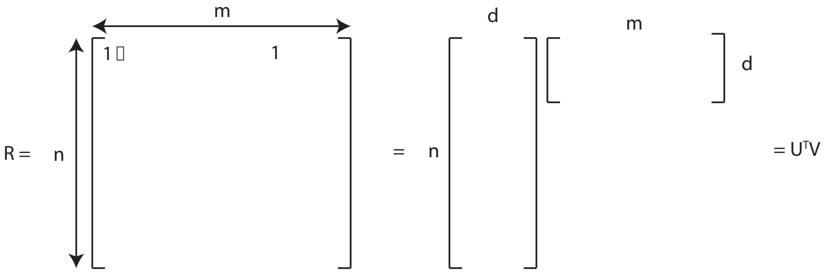
$$J(x^{(1)}, \ldots, x^{(m)}) = \sum_{u=1}^{n}\sum_{v=1}^{m}\delta_{uv}(R_{uv} - (\theta^{(u)})^T x^{(v)})^2 + \lambda\sum_{v=1}^{m}(x^{(v)})^T x^{(v)} + \lambda\sum_{u=1}^{n}(\theta^{(u)})^T \theta^{(u)}$$

- Repeat:
  Given $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(n)}$, select $x^{(1)}, x^{(2)}, \ldots, x^{(m)}$ to minimise J
  Given $x^{(1)}, x^{(2)}, \ldots, x^{(m)}$, select $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(n)}$ to minimise J
- Each update requires solving a least squares problem: use gradient descent or closed-form solution. This is called the **alternating least-squares** algorithm
- Recommendation: predicted rating by user u of item v is $(\theta^{(u)})^T x^{(v)}$

Matrix Factorisation

Another way to think about the same thing ...

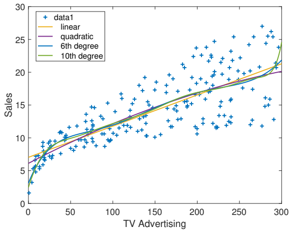
- Observe ratings R_{uv} by user u for item v. Gather these into ratings matrix R. We want to predict the missing entries in R.
- To proceed, assume R is low rank $d \ll n, m$ ....

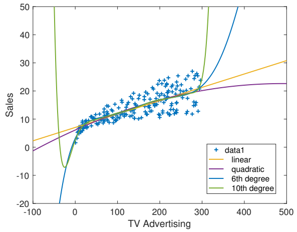


- Hypothesis: $R = U^T V$, but the elements of U and V are unknown.
- Cost Function: $\frac{1}{m}\sum_{u,v}(R_{uv} - (U^T V)_{uv})^2 + \lambda U^T U + \lambda V^T V$

Some Issues ...

- Cold-start (new user, new item)
- Popularity bias: hard to recommend to someone with unique tastes
    - Good quality data is always a key issue. Even with lots of data our model may not generalise well i.e. predict well for data outside the training set.



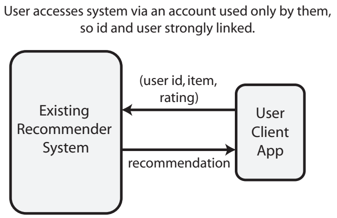


- What is the intrinsic noise when making predictions anyway ? E.g. For Netflix data set the state of the art is RMSE of about 0.9. Ratings are concentrated between 3 and 5. So 4 ± 0.9 covers almost the whole range.

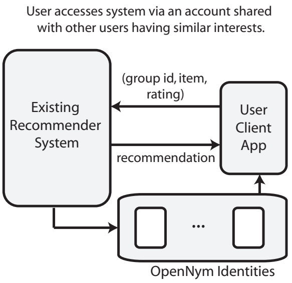
Issues

- Shilling attacks/adversarial data
  - Create costly barrier to keep bots etc out e.g. booking.com requires paying for a room in hotel before a review can be submitted.
  - Create barrier by building reputation over time e.g. stackoverflow
- And then there's the question of privacy ...
  ... US, Europe and Asia have very different privacy regulations.
  - As access control (couched as "consent") ...
  - Adding noise/perturbing the data (k-anonymity, differential privacy etc). Privacy comes at the cost of poorer performance.
  - Hiding in the crowd ?

Privacy by Design: Personalised Recommendations[1]


User accesses system via an account used only by them,
so id and user strongly linked.
Existing Recommender System
(user id, item, rating)
recommendation
User Client App

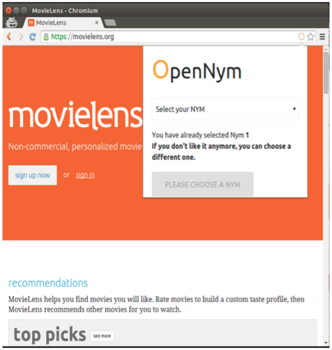
Conventional


User accesses system via an account shared with other users having similar interests.
Existing Recommender System
(group id, item, rating)
recommendation
User Client App
...
OpenNym Identities

OpenNym

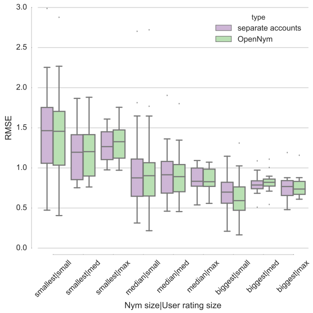
[1]Checco,A.,Bianchi,G.,DL,2017, BLC: Private Matrix Factorization Recommenders via Automatic Group Learning. ACM Trans Security and Privacy, 20(2).

Privacy by Design: Personalised Recommendations



You have already selected Nym 1
If you don't like it anymore, you can choose a different one.

MovieLens helps you find movies you will like. Rate movies to build a custom taste profile, then MovieLens recommends other movies for you to watch.



Privacy by Design: Personalised Recommendations

Summary of the RMSE performance using validation sets from [1][2].

Dataset	BMF	ALSWR	SVD++	SGD	Bias SVD	BLC	BLC local	(nyms)
Jester	4.33	5.64	5.54	5.72	5.82	4.30	**4.20**	64
Movielens	0.85	1.51	1.42	1.24	1.23	0.87	**0.83**	26
Dating	1.93	4.72	4.68	5.17	3.96	1.91	**1.88**	14
Books	1.94	4.71	4.73	5.18	3.95	1.96	**1.87**	1
Netflix	0.95	1.56	1.54	1.29	1.38	0.98	**0.97**	128

[2] [1] R. Kannan, M. Ishteva, H. Park. Bounded matrix factorization for recommender system. Knowledge and Information Systems 39, 3 (2014), 491511.