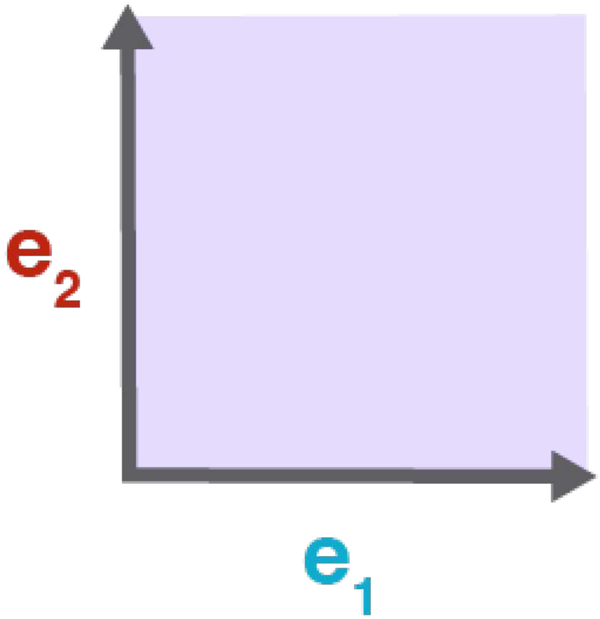
Applicazioni Informatiche del Machine Learning

04 - Matrix Factorization and Recommender Systems

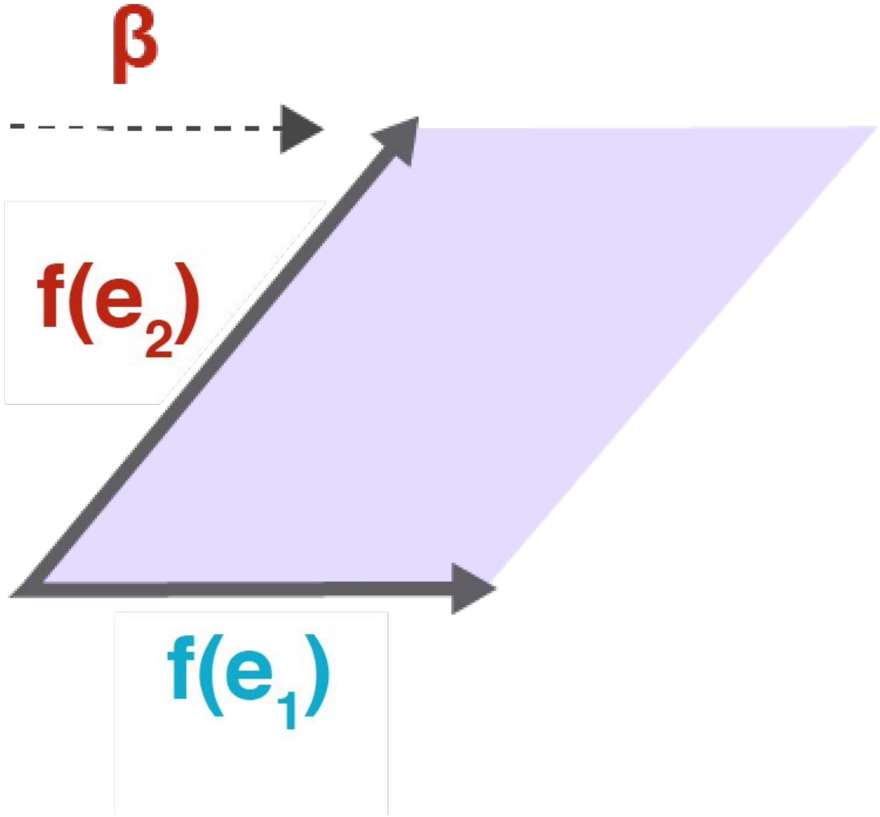
Sapienza Università di Roma

Fabrizio Silvestri

What’s a Matrix?



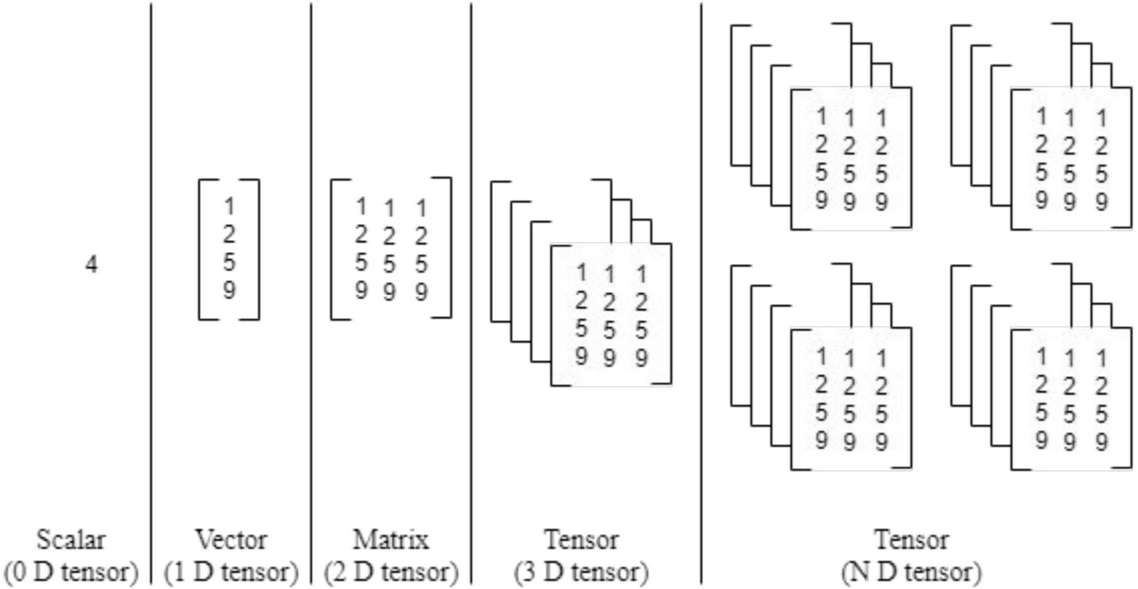
e2
e1



β
f(e2)
f(e1)

Data Organization

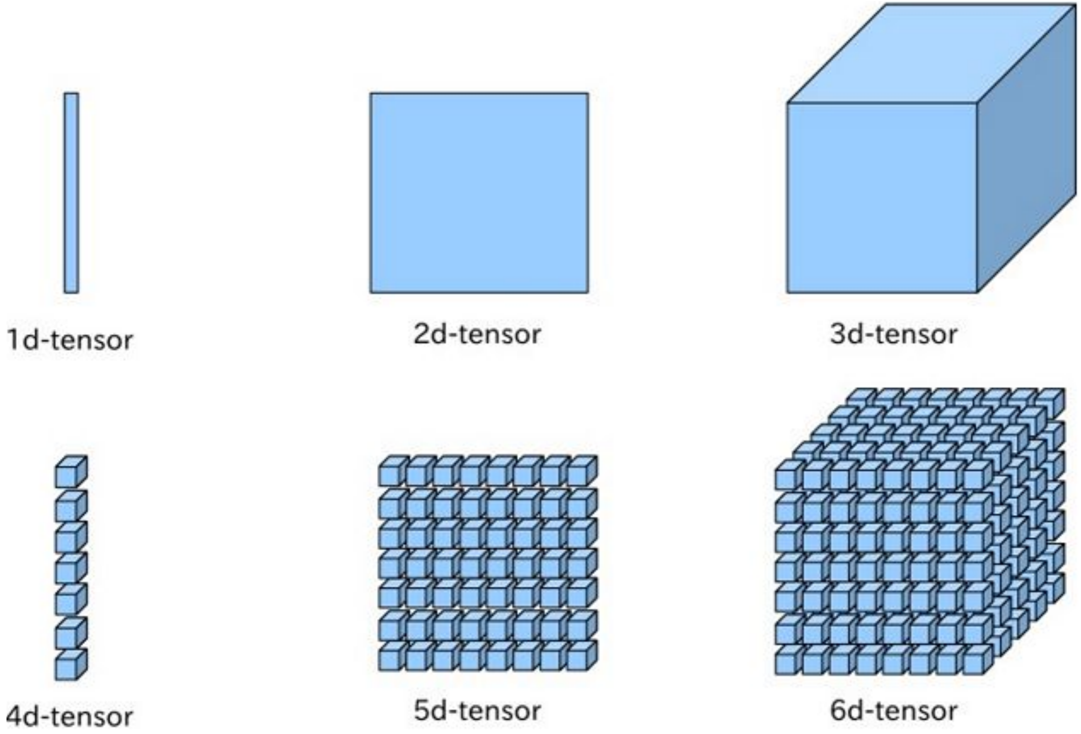
- We assume that x and y are ***tensors***



SAPIENZA UNIVERSITÀ DI ROMA

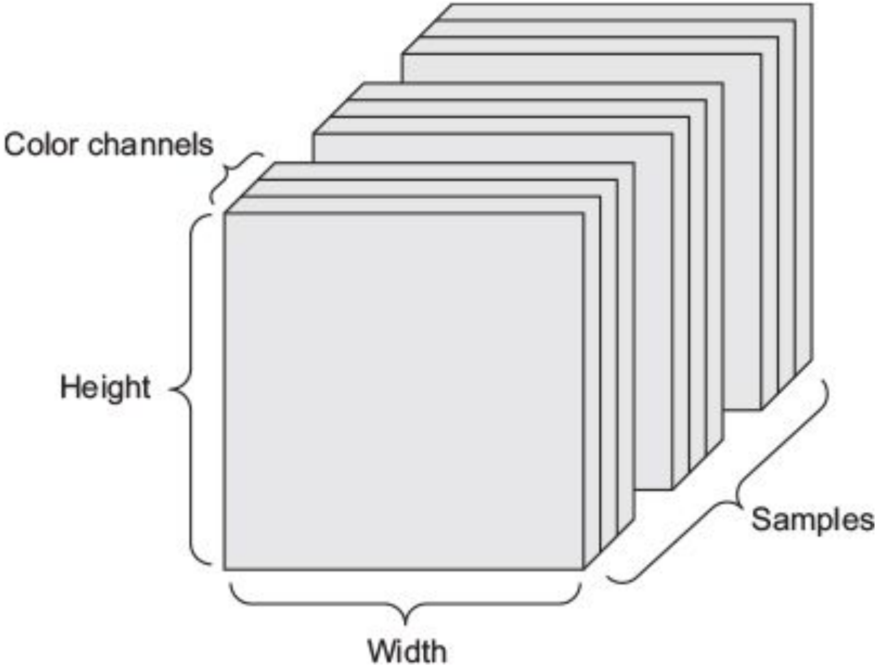
Data Organization

- We assume that x and y are ***tensors***



Data Organization

- We assume that x and y are ***tensors***



Why a Matrix (Tensor) to Represent Data?

- For regression and classification of factored tabular data, gradient boosting, sometimes called gradient boosting machines (GBM) or gradient boosted regression trees (GBRT), has become a very popular method.
- As the name implies, gradient boosting is a form of boosting using gradient descent.
- In gradient boosting we add new boosting hypotheses, which pay attention not to specific examples, but to the gradient between the right answers and the answers given by the previous hypotheses.

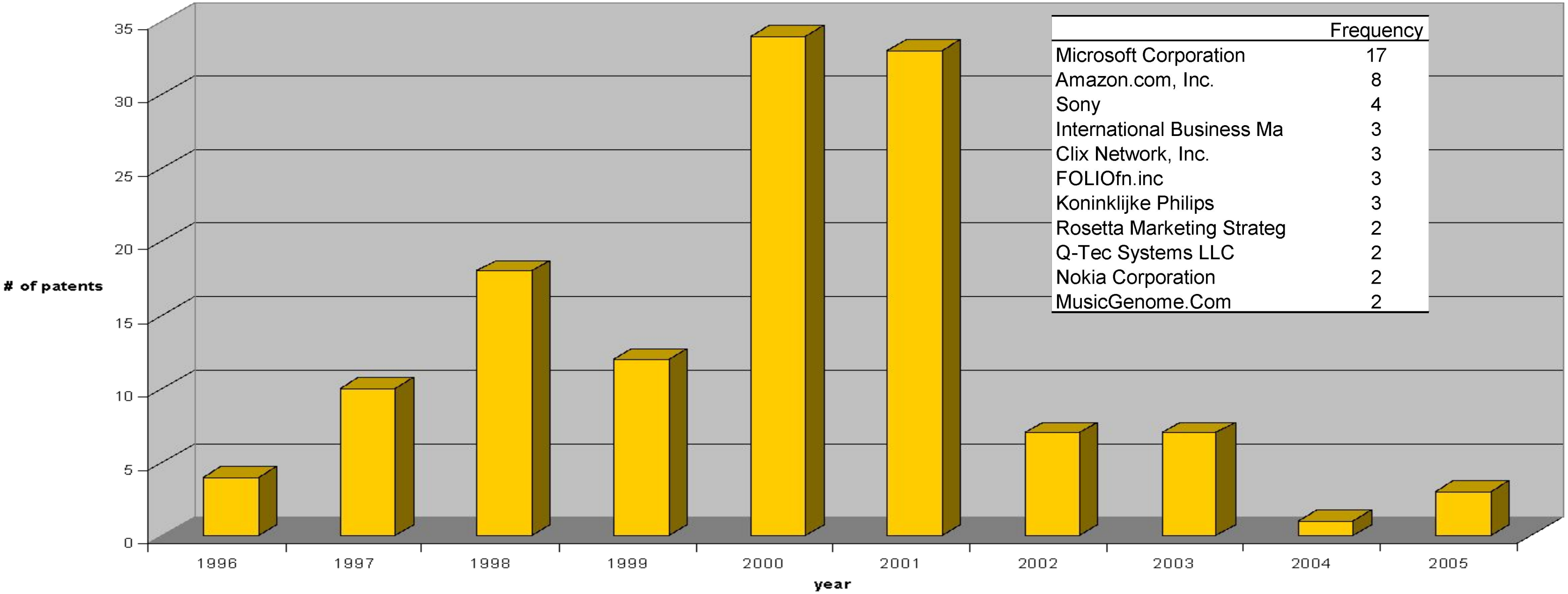
Recommender Systems

Collaborative Filtering

- **Recommendation systems** make predictions of items of interest based on user information and/or product characteristics
- **Collaborative filtering systems** make predictions what items interest a user by using information from other users.
- **Origin:** Information Tapestry project at Xerox PARC.
- *System-Input:*
  - Active: ratings by users, text comments, expert opinions
  - Passive: purchase data, usage data, browsing data
- *Taxonomy:*
  - attribute based (this author also wrote …)
  - item-to-item (people who bought this item also bought …)
  - people-to-people (users like you …)
- *Method:*
  - Memory-based: use past data and matching heuristics
  - Model-based: use models to make predictions

SAPIENZA
UNIVERSITÀ DI ROMA

Patents Filed 1995-2005



	Frequency
Microsoft Corporation	17
Amazon.com, Inc.	8
Sony	4
International Business Ma	3
Clix Network, Inc.	3
FOLIOfn.inc	3
Koninklijke Philips	3
Rosetta Marketing Strateg	2
Q-Tec Systems LLC	2
Nokia Corporation	2
MusicGenome.Com	2

SAPIENZA
UNIVERSITÀ DI ROMA

Developments in Practice

- **Massive Data:**
  - Amazon: over 6 billion product reviews
  - TiVo: 100 million ratings of 30,000 TV shows
  - Google News: millions of news items from 4500 sources updated minute-by-minute
- **Shifts:**
  - from collaborative filtering to hybrid systems
  - from ratings data to purchase/usage data
  - from e-tailer systems to stand-alone services
  - to integration with social network sites

Usually Ratings

Libri › Letteratura e narrativa › Racconti e antologie

Il problema dei tre corpi Copertina rigida – 28 marzo 2023

di Cixin Liu (Autore), Benedetta Tavani (Traduttore)

4,5 ★★★★☆ 614 voti

Cofanetti dalla serie Il problema dei tre corpi

Visualizza tutti i formati ed edizioni

Questo articolo è acquistabile con Carta Cultura Giovani, Carta del Merito e/o Carta del Docente quando venduto e spedito da Amazon: Sono esclusi prodotti di Venditori terzi del Marketplace. Il Bonus è strettamente personale e può essere utilizzato esclusivamente dal suo titolare. Termini e condizioni

Opzioni di acquisto e componenti aggiuntivi

Pagamenti rateali

A partire da 8,87 €/mese (3 mesi, senza interessi)

Un progetto militare segreto nella Cina della Rivoluzione Culturale. Un messaggio inviato nello spazio. Un mondo alieno destinato a sprofondare nel Caos. Forse l'inizio di una nuova era della storia umana.

Segnala un problema con questo prodotto

Data di

Some Problems with Ratings

- **Cold Start.** Before an individual has interacted with the recommendation system, no information is available that enables the system to generate useful recommendations. That makes these systems unsuitable for customer retention
- **Missingness.** Customers rate only a very small subset of all available items, perhaps only those they like or dislike and the ratings history of any particular customer is extremely sparse. In addition, the product rating data is missing non-randomly (Ying, Feinberg and Wedel 2006).
- **Scale Usage.** Many recommendation systems ask customers to award products 1-5 stars. But, people use scales differently. Recommendations based on ratings may reflect scale usage behavior rather than product preference (Rossi, Gilula and Allenby 2001).
- **Shilling.** Users (human or agent) may provide specially crafted ratings that cause the recommendation system to make the desired recommendations. Shilling attacks have been shown to be effective in particular for infrequently recommended items (Lam and Riedl 2004).
- **Endogeneity.** Choice behavior from customers is constrained by the recommendations based on purchase/usage received in the past. For model-based approaches biases will accumulate and the quality of the recommendation will decline (Ebbes, Wedel, Bockenholt and Steerneman 2005).
- **Scalability.** Model-based recommendation systems proposed in the academic literature are estimated with MCMC algorithms that are not scalable to datasets with the number of individuals and attributes encountered in practice (Ridgeway and Madigan 2002).

Some Problems with Ratings

- **Cold Start.** Before an individual has interacted with the recommendation system, no information is available that enables the system to generate useful recommendations. That makes these systems unsuitable for customer retention
- **Missingness.** Customers rate only a very small subset of all available items, perhaps only those they like or dislike and the ratings history of any particular customer is extremely sparse. In addition, the product rating data is missing non-randomly (Ying, Feinberg and Wedel 2006).
- **Scale Usage.** Many recommendation systems ask customers to award products 1-5 stars. But, people use scales differently. Recommendations based on ratings may reflect scale usage behavior rather than product preference (Rossi, Gilula and Allenby 2001).
- **Shilling.** Users (human or agent) may provide specially crafted ratings that cause the recommendation system to make the desired recommendations. Shilling attacks have been shown to be effective in particular for infrequently recommended items (Lam and Riedl 2004).
- **Endogeneity.** Choice behavior from customers is constrained by the recommendations based on purchase/usage received in the past. For model-based approaches biases will accumulate and the quality of the recommendation will decline (Ebbes, Wedel, Bockenholt and Steerneman 2005).
- **Scalability.** Model-based recommendation systems proposed in the academic literature are estimated with MCMC algorithms that are not scalable to datasets with the number of individuals and attributes encountered in practice (Ridgeway and Madigan 2002).

Some Problems with Ratings

- **Cold Start.** Before an individual has interacted with the recommendation system, no information is available that enables the system to generate useful recommendations. That makes these systems unsuitable for customer retention
- **Missingness.** Customers rate only a very small subset of all available items, perhaps only those they like or dislike and the ratings history of any particular customer is extremely sparse. In addition, the product rating data is missing non-randomly (Ying, Feinberg and Wedel 2006).
- **Scale Usage.** Many recommendation systems ask customers to award products 1-5 stars. But, people use scales differently. Recommendations based on ratings may reflect scale usage behavior rather than product preference (Rossi, Gilula and Allenby 2001).
- **Shilling.** Users (human or agent) may provide specially crafted ratings that cause the recommendation system to make the desired recommendations. Shilling attacks have been shown to be effective in particular for infrequently recommended items (Lam and Riedl 2004).
- **Endogeneity.** Choice behavior from customers is constrained by the recommendations based on purchase/usage received in the past. For model-based approaches biases will accumulate and the quality of the recommendation will decline (Ebbes, Wedel, Bockenholt and Steerneman 2005).
- **Scalability.** Model-based recommendation systems proposed in the academic literature are estimated with MCMC algorithms that are not scalable to datasets with the number of individuals and attributes encountered in practice (Ridgeway and Madigan 2002).

Some Problems with Ratings

- **Cold Start.** Before an individual has interacted with the recommendation system, no information is available that enables the system to generate useful recommendations. That makes these systems unsuitable for customer retention
- **Missingness.** Customers rate only a very small subset of all available items, perhaps only those they like or dislike and the ratings history of any particular customer is extremely sparse. In addition, the product rating data is missing non-randomly (Ying, Feinberg and Wedel 2006).
- **Scale Usage.** Many recommendation systems ask customers to award products 1-5 stars. But, people use scales differently. Recommendations based on ratings may reflect scale usage behavior rather than product preference (Rossi, Gilula and Allenby 2001).
- **Shilling.** Users (human or agent) may provide specially crafted ratings that cause the recommendation system to make the desired recommendations. Shilling attacks have been shown to be effective in particular for infrequently recommended items (Lam and Riedl 2004).
- **Endogeneity.** Choice behavior from customers is constrained by the recommendations based on purchase/usage received in the past. For model-based approaches biases will accumulate and the quality of the recommendation will decline (Ebbes, Wedel, Bockenholt and Steerneman 2005).
- **Scalability.** Model-based recommendation systems proposed in the academic literature are estimated with MCMC algorithms that are not scalable to datasets with the number of individuals and attributes encountered in practice (Ridgeway and Madigan 2002).

Some Problems with Ratings

- **Cold Start.** Before an individual has interacted with the recommendation system, no information is available that enables the system to generate useful recommendations. That makes these systems unsuitable for customer retention
- **Missingness.** Customers rate only a very small subset of all available items, perhaps only those they like or dislike and the ratings history of any particular customer is extremely sparse. In addition, the product rating data is missing non-randomly (Ying, Feinberg and Wedel 2006).
- **Scale Usage.** Many recommendation systems ask customers to award products 1-5 stars. But, people use scales differently. Recommendations based on ratings may reflect scale usage behavior rather than product preference (Rossi, Gilula and Allenby 2001).
- **Shilling.** Users (human or agent) may provide specially crafted ratings that cause the recommendation system to make the desired recommendations. Shilling attacks have been shown to be effective in particular for infrequently recommended items (Lam and Riedl 2004).
- **Endogeneity.** Choice behavior from customers is constrained by the recommendations based on purchase/usage received in the past. For model-based approaches biases will accumulate and the quality of the recommendation will decline (Ebbes, Wedel, Bockenholt and Steerneman 2005).
- **Scalability.** Model-based recommendation systems proposed in the academic literature are estimated with MCMC algorithms that are not scalable to datasets with the number of individuals and attributes encountered in practice (Ridgeway and Madigan 2002).

Some Problems with Ratings

- **Cold Start.** Before an individual has interacted with the recommendation system, no information is available that enables the system to generate useful recommendations. That makes these systems unsuitable for customer retention
- **Missingness.** Customers rate only a very small subset of all available items, perhaps only those they like or dislike and the ratings history of any particular customer is extremely sparse. In addition, the product rating data is missing non-randomly (Ying, Feinberg and Wedel 2006).
- **Scale Usage.** Many recommendation systems ask customers to award products 1-5 stars. But, people use scales differently. Recommendations based on ratings may reflect scale usage behavior rather than product preference (Rossi, Gilula and Allenby 2001).
- **Shilling.** Users (human or agent) may provide specially crafted ratings that cause the recommendation system to make the desired recommendations. Shilling attacks have been shown to be effective in particular for infrequently recommended items (Lam and Riedl 2004).
- **Endogeneity.** Choice behavior from customers is constrained by the recommendations based on purchase/usage received in the past. For model-based approaches biases will accumulate and the quality of the recommendation will decline (Ebbes, Wedel, Bockenholt and Steerneman 2005).
- **Scalability.** Model-based recommendation systems proposed in the academic literature are estimated with MCMC algorithms that are not scalable to datasets with the number of individuals and attributes encountered in practice (Ridgeway and Madigan 2002).

Studies have shown that

- Recommendation agents may reduce the prices paid (Diehl, Kornish, and Lynch 2003) and improve decision quality and efficiency (Ariely, Lynch, and Aparicio 2004; Haübl and Trifts 2000; West 1996), and may influence user opinions (Cosley e.a. 2003; Haubel & Murray 2003). Agents and collaborative filtering learn at different rates (Ariely, Lynch & Aparicio 2004) and their effectiveness depends on the similarity with the users (Aksoy e.a. 2006).
- **Model-based methods,** including
  - Bayes net (Breese, Heckerman, & Kadie 1998), Nearest Neighbor (Herlocker, Konstan & Riedl 2002), Tree-based (Breese, Heckerman & Kadie, 1998), Mixture (Chien & George 1999), Dual Mixture (Bodapati 2007) HB models (Ansari, Essegaier & Kohli 2000), HB selection models (Ying, Feinberg & Wedel 2004).
- in most cases show **substantial improvements** in the quality of recommendations on test datasets.
- However, the models in the academic literature are mostly estimated with MCMC algorithms and are **not scalable**.

SAPIENZA
UNIVERSITÀ DI ROMA

Collaborative Systems as a Linear Algebra Problem

[Matrix Factorization Techniques For Recommender Systems]

SAPIENZA
UNIVERSITÀ DI ROMA

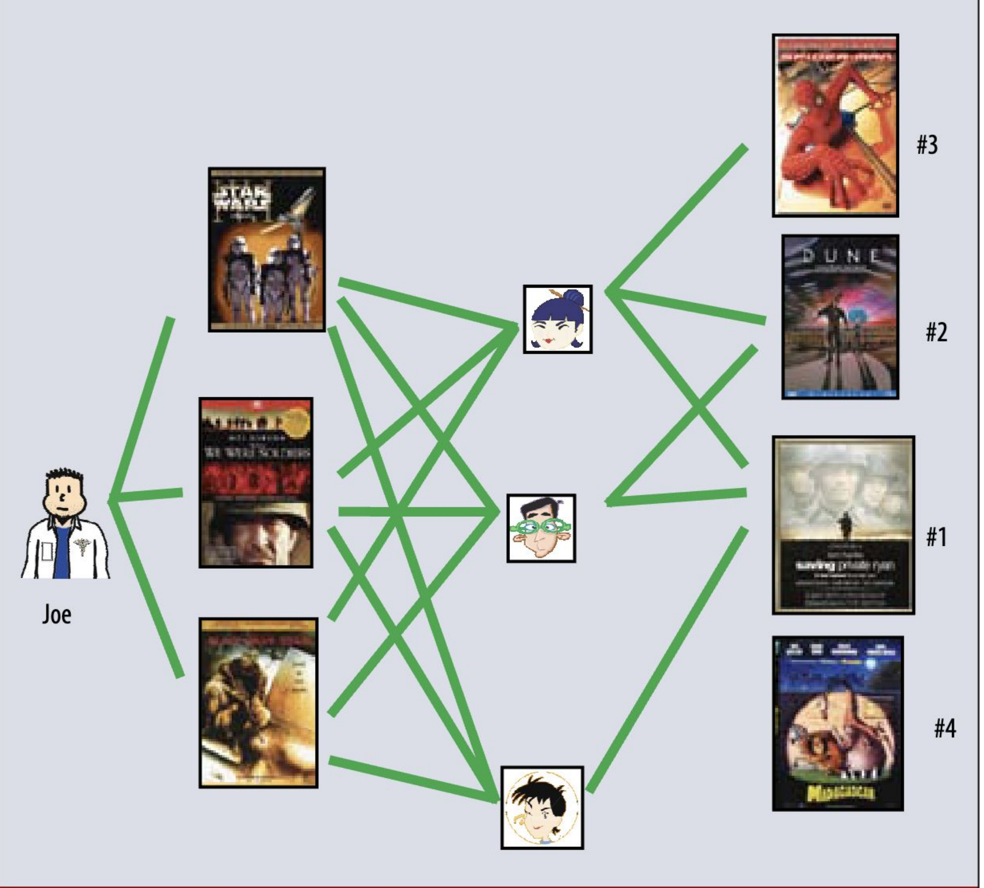


Figure 1. The user-oriented neighborhood method. Joe likes the three movies on the left. To make a prediction for him, the system finds similar users who also liked those movies, and then determines which other movies they liked. In this case, all three liked *Saving Private Ryan*, so that is the first recommendation. Two of them liked *Dune*, so that is next, and so on.

SAPIENZA
UNIVERSITÀ DI ROMA

Users’ Ratings

Libri › Letteratura e narrativa › Racconti e antologie

Il problema dei tre corpi Copertina rigida – 28 marzo 2023

di Cixin Liu (Autore), Benedetta Tavani (Traduttore)

4,5 614 voti

Cofanetti dalla serie Il problema dei tre corpi

Visualizza tutti i formati ed edizioni

Questo articolo è acquistabile con Carta Cultura Giovani, Carta del Merito e/o Carta del Docente quando venduto e spedito da Amazon: Sono esclusi prodotti di Venditori terzi del Marketplace. Il Bonus è strettamente personale e può essere utilizzato esclusivamente dal suo titolare. Termini e condizioni

Opzioni di acquisto e componenti aggiuntivi

Pagamenti rateali

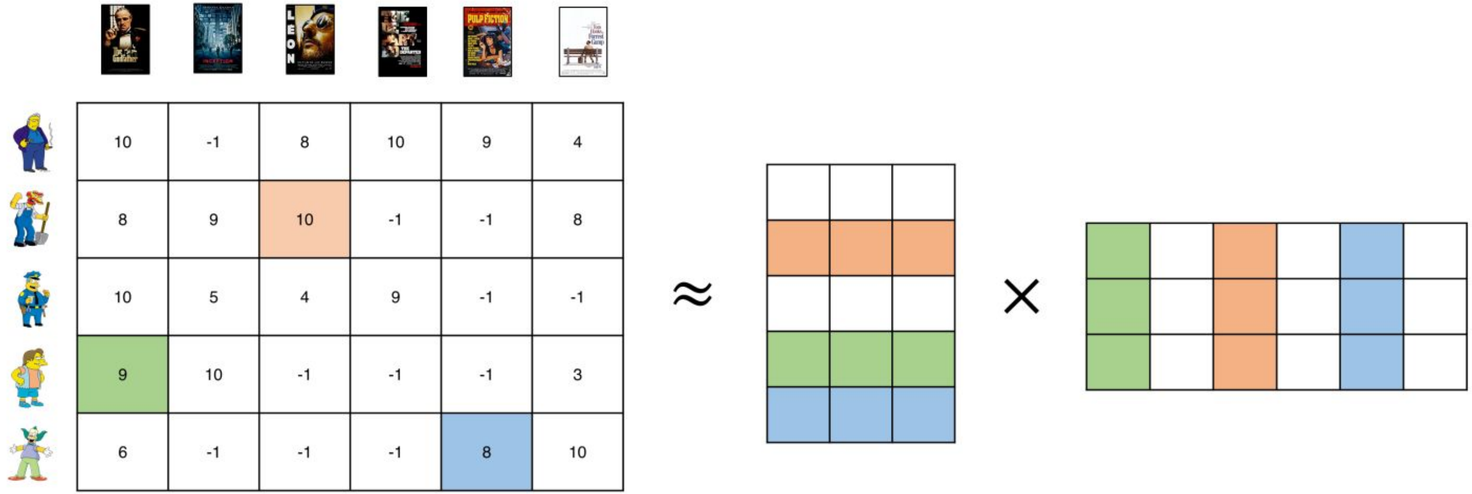
A partire da 8,87 €/mese (3 mesi, senza interessi)

Un progetto militare segreto nella Cina della Rivoluzione Culturale. Un messaggio inviato nello spazio. Un mondo alieno destinato a sprofondare nel Caos. Forse l'inizio di una nuova era della storia umana.

Segnala un problema con questo prodotto

Data di

What’s Matrix Factorization?



10	-1	8	10	9	4
8	9	10	-1	-1	8
10	5	4	9	-1	-1
9	10	-1	-1	-1	3
6	-1	-1	-1	8	10

$\approx$ $\times$

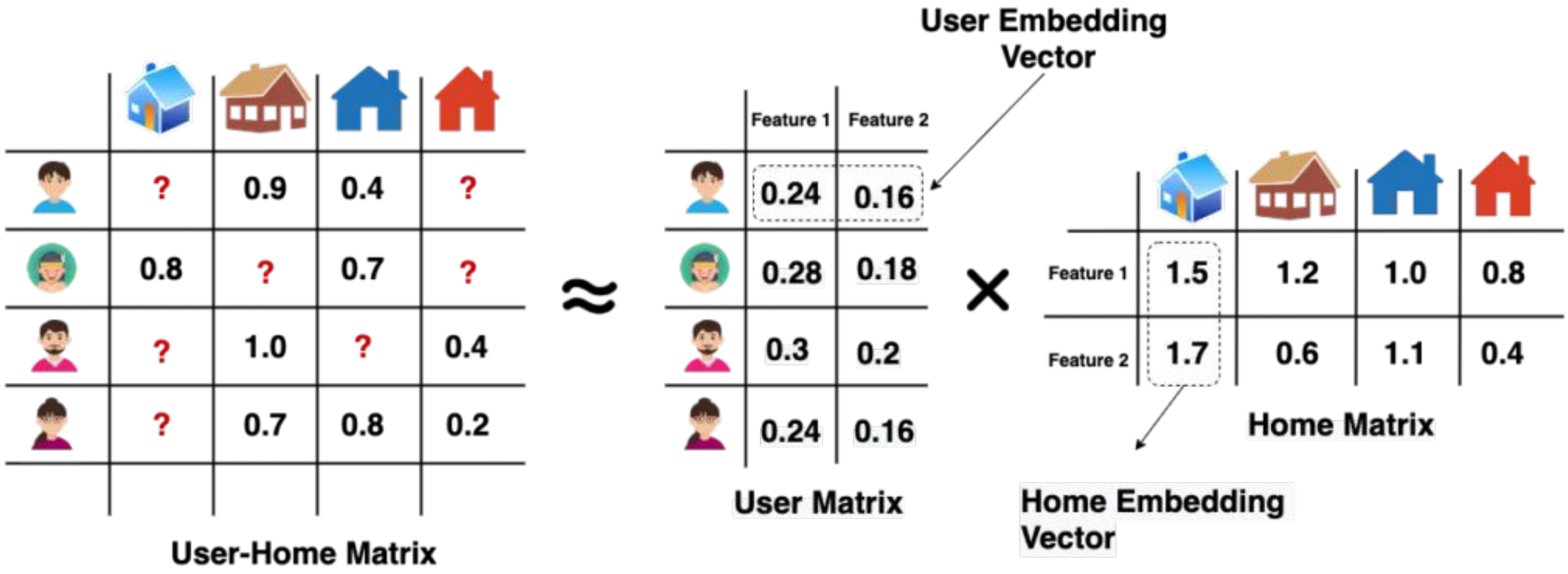
User-item Interaction Matrix (R)

User Matrix (Q)

Item Matrix (P)

SAPIENZA
UNIVERSITÀ DI ROMA

What’s Matrix Factorization?



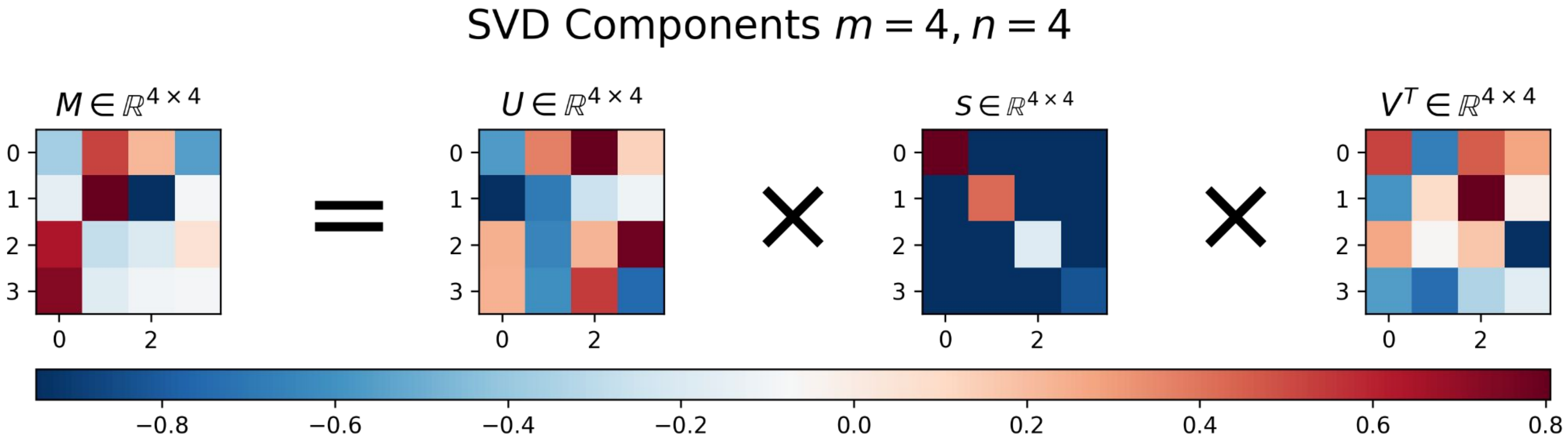
Basic Matrix Factorization Method

- Represents users and items in a **latent factor space**.

$$r_{ui} = p_u^T q_i$$

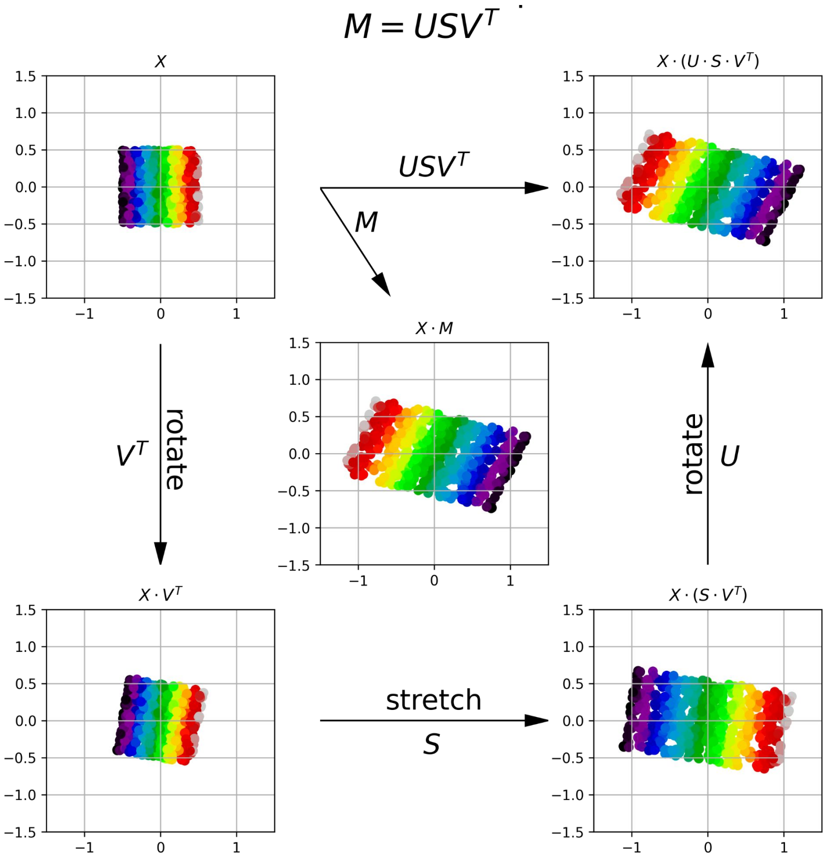
- Predict user (u) ratings (r) for items (i).
- p == latent user factor
- q == latent item factor

Singular Value Decomposition



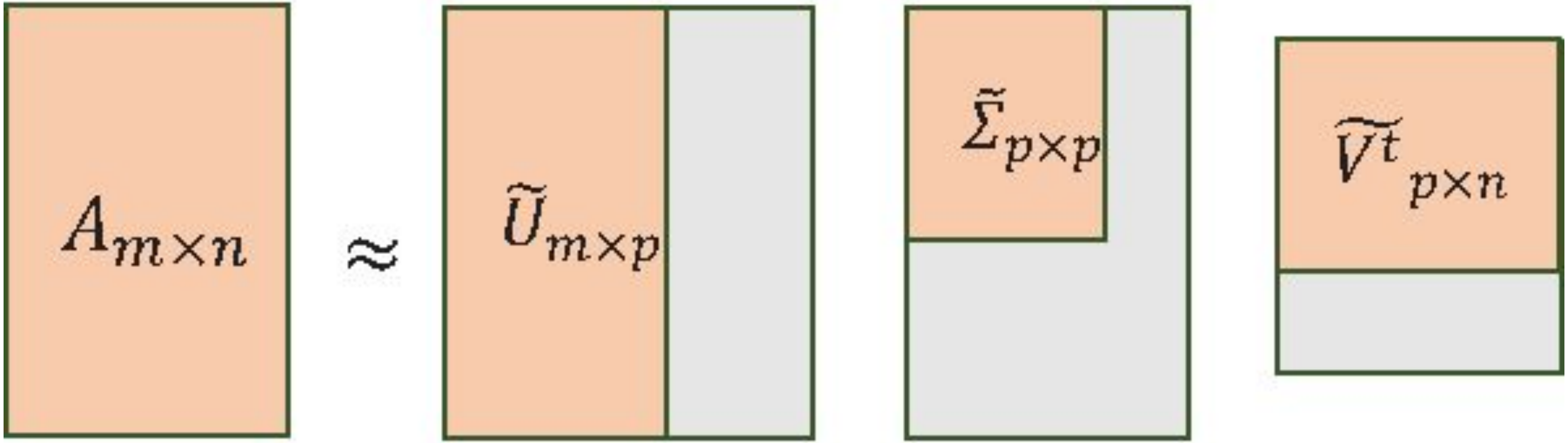
SAPIENZA
UNIVERSITÀ DI ROMA

Breakdown of SVD Operations



Sapienza Università di Roma

Truncated SVD



A Simple RecSys with SVD

A Detour: Gradient Descent

Gradient Descent

- Imagine you want to minimize a function L[**ϕ**]
- Starts with initial parameters **ϕ** = $[\phi_0, \phi_1, \ldots, \phi_N]^T$ and iterates two steps:

Step 1. Compute the derivatives of the loss with respect to the parameters:

$$\mathrm{L}[\phi] = \sum_{i=1}^{I} \ell_i \qquad \frac{\partial L}{\partial \phi} = \begin{bmatrix} \frac{\partial L}{\partial \phi_0} \\ \frac{\partial L}{\partial \phi_1} \\ \vdots \\ \frac{\partial L}{\partial \phi_N} \end{bmatrix}.$$

Step 2. Update the parameters according to the rule:

Learning Rate

$$\phi \longleftarrow \phi - \boxed{\alpha} \frac{\partial L}{\partial \phi},$$

where the positive scalar α determines the magnitude of the change.

SAPIENZA UNIVERSITÀ DI ROMA

Gradient Descent: Linear Regression Example

- The model y = f[x,$\boldsymbol{\phi}$] maps a scalar input x to a scalar output y and has parameters $\boldsymbol{\phi}$ = [ϕ_0, ϕ_1]^T, which represent the y-intercept and the slope:

$$
\begin{aligned}
y &= \mathrm{f}[x, \boldsymbol{\phi}] \\
&= \phi_0 + \phi_1 x
\end{aligned}
$$

- Given a dataset {x_i, y_i} containing I input/outp ast squares loss function:

$\ell_i = (\phi_0 + \phi_1 x_i - y_i)^2$ is the individual contribution to the loss from the ith training example

$$
\begin{aligned}
\mathrm{L}[\boldsymbol{\phi}] &= \sum_{i=1}^{I} \boxed{\ell_i} = \sum_{i=1}^{I} (\mathrm{f}[x_i, \boldsymbol{\phi}] - y_i)^2 \\
&= \sum_{i=1}^{I} (\phi_0 + \phi_1 x_i - y_i)^2
\end{aligned}
$$

SAPIENZA
UNIVERSITÀ DI ROMA

Gradient Descent: Linear Regression Example

- The derivative of the loss function with respect to the parameters can be decomposed into the sum of the derivatives of the individual contributions:

$$\frac{\partial L}{\partial \phi} = \frac{\partial}{\partial \phi} \sum_{i=1}^{I} \ell_i = \sum_{i=1}^{I} \frac{\partial \ell_i}{\partial \phi}$$

$$\ell_i = (\phi_0 + \phi_1 x_i - y_i)^2$$

- where these are given by (take 5 mins to compute it):

$$\frac{\partial \ell_i}{\partial \phi} = \begin{bmatrix} \frac{\partial \ell_i}{\partial \phi_0} \\ \frac{\partial \ell_i}{\partial \phi_1} \end{bmatrix} = \begin{bmatrix} 2(\phi_0 + \phi_1 x_i - y_i) \\ 2x_i(\phi_0 + \phi_1 x_i - y_i) \end{bmatrix}$$

SAPIENZA UNIVERSITÀ DI ROMA

Gradient Descent: Linear Regression Example

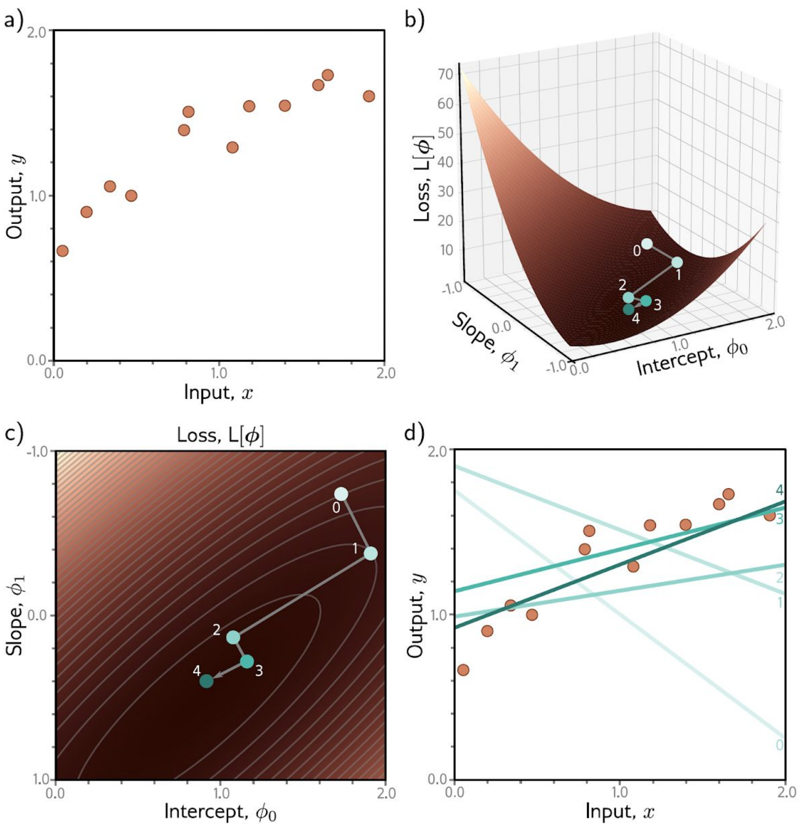


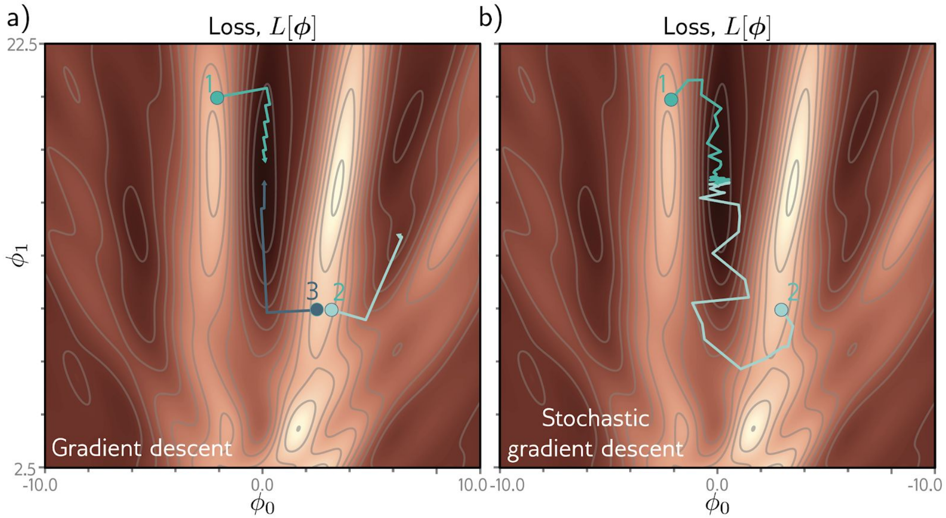
Figure 6.1 Gradient descent for the linear regression model. a) Training set of $I = 12$ input/output pairs $\{x_i, y_i\}$. b) Loss function showing iterations of gradient descent. We start at point 0 and move in the steepest downhill direction until we can improve no further to arrive at point 1. We then repeat this procedure. We measure the gradient at point 1 and move downhill to point 2 and so on. c) This can be visualized better as a heatmap, where the brightness represents the loss. After only four iterations, we are already close to the minimum. d) The model with the parameters at point 0 (lightest line) describes the data very badly, but each successive iteration improves the fit. The model with the parameters at point 4 (darkest line) is already a reasonable description of the training data.

Colab time: Gradient Descent

[Gradient_Descent]

Stochastic Gradient Descent

- Starts with initial parameters $\boldsymbol{\phi} = [\phi_0, \phi_1, \ldots, \phi_N]^T$ and iterates the same two steps of Gradient Descent, except…
- … The gradient is computed for each element (x_i, y_i) in the dataset: $\frac{\partial \ell_i}{\partial \phi}$



Regularized Matrix Factorization

- Minimizes regularized squared error on known ratings.

$$min_{p,q} \sum (r_{ui} - p_u^T q_i)^2 + \lambda(||p_u||^2 + ||q_i||^2)$$

- Min is found by Gradient Descent
- Colab time: Recommender System with Matrix Factorization

Nonnegative Matrix Factorization

- Minimizes regularized squared error on known ratings.

$$min_{p,q} \sum (r_{ui} - p_u^T q_i)^2$$

- Min is found by Gradient Descent
- P, and Q must contain elements that are nonnegative
- Colab time: [Recommender System with Matrix Factorization using Scikit Learn]

Applicazioni Informatiche del Machine Learning

End of Lecture

04 - Matrix Factorization and Recommender Systems

SAPIENZA
UNIVERSITÀ DI ROMA

Fabrizio Silvestri