Chapter 3: Matrix Factorization

Making recommendations by discovering latent features

Let's say we work in a vinyl record and CD store downtown. A customer comes in and asks us for a recommendation. We ask the obvious question "What do you like?" and she responds with

There are several ways we might come up with a recommendation for her. One method we used in previous chapters is to reflect on regular customers to our store who bought albums from those artists and think what else they bought. Perhaps we notice that recently, people who bought Taylor Swift CDs also bought CDs by Miranda Lambert and we recommend Miranda Lambert to the new customer. This is a two step process. First, we determine previous customers who are most similar in their buying behavior to the person standing in front of us and second, we look at what those previous customers bought and then use that information to make recommendations to our current customer. Again, this is a method we have been using in the previous two chapters.

Another way we might come up with a recommendation is as follows. We know Taylor Swift and Carrie Underwood CDs share certain features. They both, obviously, have prominent female vocals. They both feature singer-songwriters. They both have country influences and no PBR&B influences.

The term PBR&B, aka hipster R&B and R neg B, is a portmanteau of PBR–Pabst Blue Ribbon, the hipster beer of choice–and R&B

Then we think "Hey, Miranda Lambert CDs also have prominent female vocals and have country influences but no PBR&B influences, and we recommend Miranda Lambert to our new customer. With this recommendation method we extract a set of features from CDs this person likes and then think what other CDs share these features.

Let's check this out a bit further. We will restrict ourselves to two features (country and PBR&B influences), six artists (Taylor Swift, Miranda Lambert, Carrie Underwood, Jhené Aiko, Kelela and Tinashe) and two customers (Sarah and Jake). As owners of the vinyl record store we have gone through and meticulously rated artists on these features.

Artist	Country	PBR&B
Taylor Swift	0.90	0.05
Miranda Lambert	0.98	0.00
Carrie Underwood	0.95	0.03
Jhene Aiko	0.01	0.99
Kelela	0.10	0.98
Tinashe	0.05	0.96

So Taylor Swift exudes a lot of country influences (0.90) but little PBR&B (0.05).

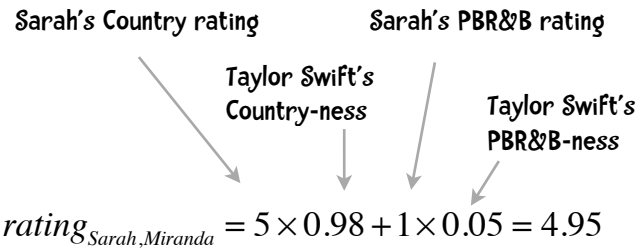
When customers come into our store we ask them on a scale of 0 to 5 how well they like country and how well they like PBR&B:

	Country	PBR&B
Sarah	5	1
Jake	0	5

Suppose Sarah comes into the store, has never heard of Miranda Lambert, and we are trying to predict how she might rate her. She rated country music a 5 and Miranda is 0.98 country so we multiply those numbers together:

$$5 \times 0.98 = 4.9$$

We do the same for the PBR&B numbers and add them together to get our estimate of Sarah's rating of Miranda Lambert.



$$rating_{Sarah,Miranda} = 5 \times 0.98 + 1 \times 0.05 = 4.95$$

sharpen your pencil

What are Jake's ratings of Taylor Swift and Jhené Aiko?

sharpen your pencil - answer

What are Jake's ratings of Taylor Swift and Jhené Aiko?

$$rating_{Jake,TaylorSwift} = 0 \times 0.9 + 5 \times 0.05 = 0.25$$

$$rating_{Jake,JhenéAiko} = 0 \times 0.01 + 5 \times 0.00 = 4.95$$

A reflection ...

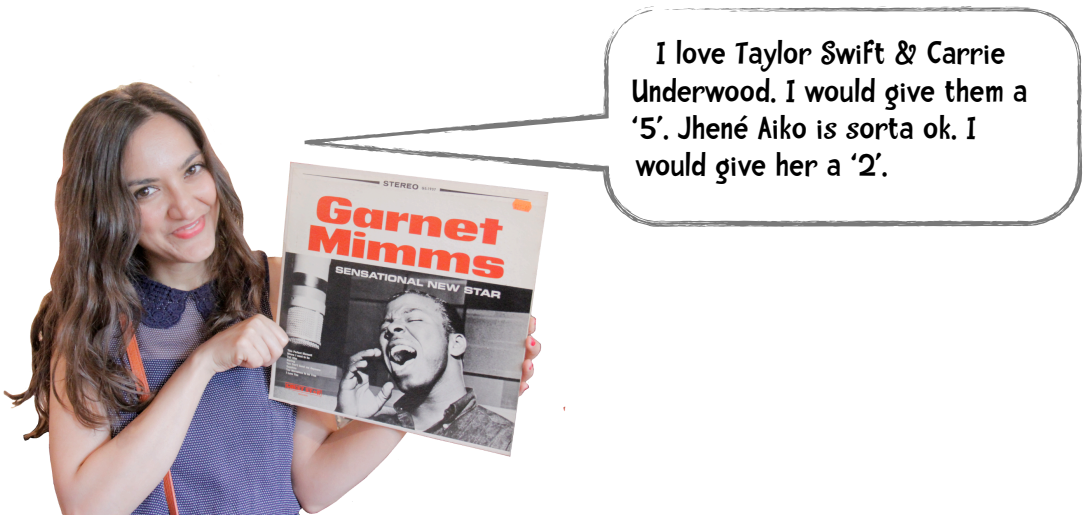
Okay. Let's reflect on the two methods we used.

1. We asked our customer, Sarah, what musical artists she likes. She said Taylor Swift and Carrie Underwood. Then we thought, "what other customers like Taylor Swift and Carrie Underwood and what additional artists do those customers like?" and we made our recommendation—Miranda Lambert.

2. We asked Sarah, How much she likes country music and how much she likes PBR&B. She said she really likes country and said '5' and gave PBR&B a '1'. Then we thought, "what artists are country artists" and we made our recommendation—Miranda Lambert.

Let's see if we can combine these methods and if such a combination makes sense.

We will go back to asking our customers what musical artists they like:



Now we are going to do a bit of brainwork. When Sarah says she likes Taylor Swift and Carrie Underwood, we are going to be thinking "Hey, I bet Sarah likes Country Music and in particular female country artists." Then we are going to think "What other artists are categorized as country artists?" and recommend those to Sarah. When Adam says he likes Three Days Grace, Disturbed and Faith No More, we think

Let's give this a try. We survey a number of our customers and get something like the following (a question mark indicates that that customer has not rated that artist):

Customer	Taylor Swift	Miranda Lambert	Carrie Underwood	Jhené Aiko	Kelela	Tinashe
Sarah	5	?	5	2	2	?
Miguel	2	?	?	4	5	5
Tyler	5	5	5	2	?	1
Ann	2	3	?	5	5	?
Jessica	2	1	?	5	?	?

Based on this customer artist rating table and our table of the country and PBR&B influences of various artists

Artist	Country	PBR&B
Taylor Swift	0.90	0.05
Miranda Lambert	0.98	0.00
Carrie Underwood	0.95	0.03
Jhene Aiko	0.01	0.99
Kelela	0.10	0.98
Tinashe	0.05	0.96

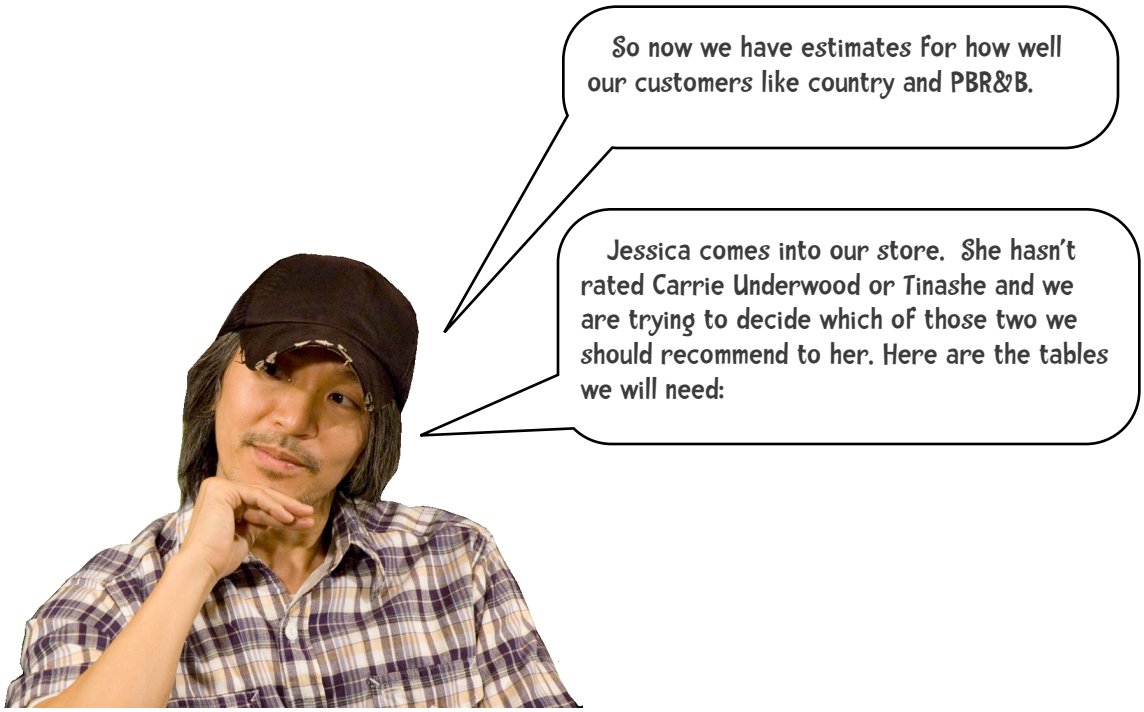
and our we would like to predict how well our customers like country and PBR&B:

Customer	Country	PBR&B
Sarah	?	?
Miguel	?	?
Tyler	?	?
Ann	?	?
Jessica	?	?

mental calisthenics

Can you fill in the table on the previous page with rough guesses?

My feeling is that you can guess these values very accurately. For example, you probably gave Sarah close to a 5 for country and a 2 for PBR&B and Jessica a 1.5 for country and a 5 for PBR&B. Why did we predict Sarah would give a '5' for country? Well, Sarah gave a '5' to Taylor Swift and Carrie Underwood. According to the artist chart both those artists are heavily country (about .925 on average). So it seems likely that Sarah likes country music. In the table on the following page, I roughed out how well our customers like Country and PBR&B based on the previous table of artist ratings.



Customer	Country	PBR&B
Sarah	4.75	2
Miguel	2	4.75
Tyler	5	1.5
Ann	2.5	5
Jessica	1.5	5

We got these values by combining the the table of how well our customers like different artists with the table of the Country and PBR&B influences of each artist.

For now don't worry how we exactly determined these numbers. We will shortly go into the details.

Artist	Country	PBR&B
Taylor Swift	0.90	0.05
Miranda Lambert	0.98	0.00
Carrie Underwood	0.95	0.03
Jhene Aiko	0.01	0.99
Kelela	0.10	0.98
Tinashe	0.05	0.96

sharpen your pencil

What is your prediction for how well Jessica will like Carrie Underwood and Tinashe?

$$rating_{Jessica,CarrieUnderwood} = ?$$

$$rating_{Jessica,Tinashe} = ?$$

sharpen your pencil - solution

What is your prediction for how well Jessica will like Carrie Underwood and Tinashe?

$$rating_{Jessica,CarrieUnderwood} = 1.50 \times 0.95 + 5 \times 0.03 = 1.57$$

$$rating_{Jessica,Tinashe} = 1.5 \times 0.05 + 5 \times 0.96 = 4.875$$

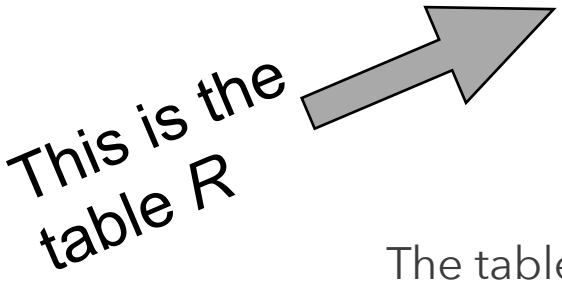
So we would recommend Tinashe to Jessica.

Let's get technical!

Let's give these tables names so we can continue talking about them throughout the chapter.

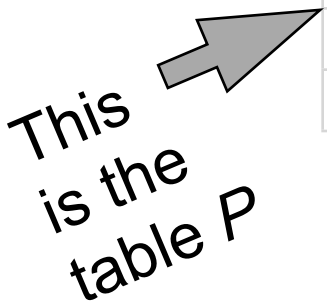
We will call the table of **r**atings users give artists R.

Customer	Taylor Swift	Miranda Lambert	Carrie Underwood	Jhené Aiko	Kelela	Tinashe
Sarah	5	?	5	2	2	?
Miguel	2	?	?	4	5	5
Tyler	5	5	5	2	?	1
Ann	2	3	?	5	5	?
Jessica	2	1	?	5	?	?



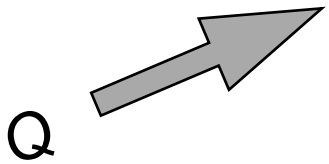
The table of how well users like various features we will call ***P***:

Customer	Country	PBR&B
Sarah	4.75	2
Miguel	2	4.75
Tyler	5	1.5
Ann	2.5	5
Jessica	1.5	5



and the table matching artists to various features we will call **Q**:

Artist	Country	PBR&B
Taylor Swift	0.90	0.05
Miranda Lambert	0.98	0.00
Carrie Underwood	0.95	0.03
Jhene Aiko	0.01	0.99
Kelela	0.10	0.98
Tinashe	0.05	0.96



We've seen that if we have *P* (the table of how well customers like particular features) and *Q* (the table matching artists to features) we can figure out *R* (how well customers like particular artists). And we have just seen that if we have *Q* and *R* we can figure out *P*.

mental calisthenics

What does "And we have just seen that if we have Q and R we can figure out P" mean? Can you restate it without using the terms P, Q, and R?

mental calisthenics

Can you match up the terms with their definitions?

P

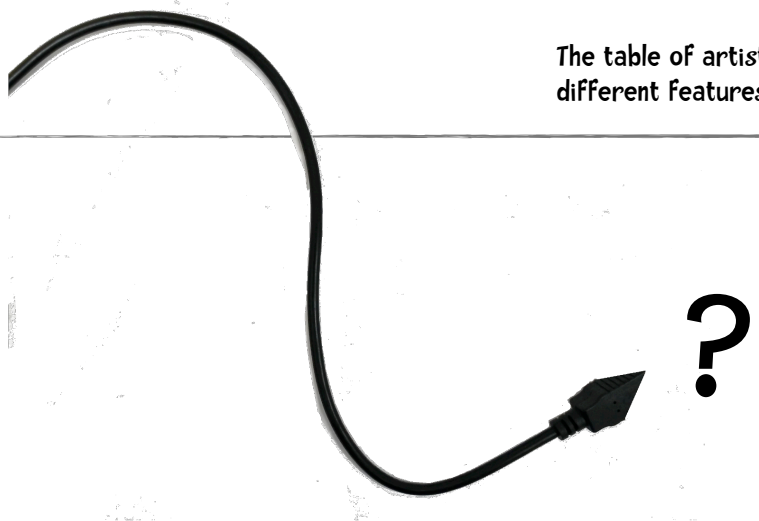
The table of customers and the ratings they give different artists.

Q

The table of customers and their ratings of different features (for ex., Country or PBR&B)

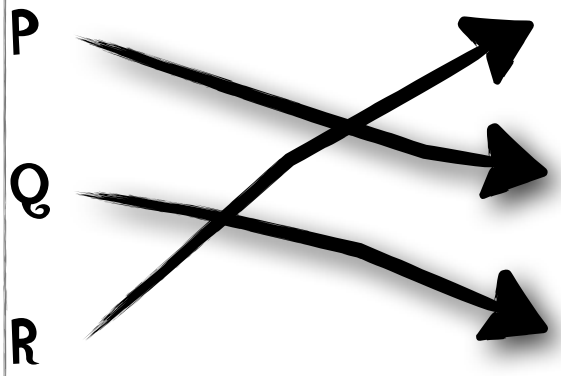
R

The table of artists and how well they match different features.



mental calisthenics

Can you match up the terms with their definitions?



The table of customers and the ratings they give different artists.

The table of customers and their ratings of different features (for ex., Country or PBR&B)

The table of artists and how well they match different features.

Now we know the meaning of *P*, *Q*, and *R*. We know we can figure out *R* given *P* and *Q* and we can figure out *P* given *R* and *Q*. But here is a question. Can we figure out Q, the table indicating the country and PBR&B influences of the artists from R the table of customers rating artists and P the table of how well customers like country and PBR&B?

What do you think?

Yes, it is the case that if we know how customers rated artists (R) and how well customers like country and PBR&B music (P) we can figure out the country and PBR&B influences of the artists (Q).

As you probably guessed from the title of this chapter, this chapter is about matrix factorization. And here we are... over ten pages into the chapter and we have not encountered the word *matrix* anywhere nor have we talked about factorization, whatever that is. Now that we covered the preliminaries, we are finally about to dive into matrix factorization.

> The good news is that we have covered all the math you need to learn (okay, maybe more like 99%). Now we are going to fancy it up by using special names and symbols. Don't be mislead into thinking we are covering something new. It is just a repeat of what we have been doing!

This would be an excellent time to take a break, get coffee, take a brisk walk or do a few yoga positions.

I think the intersection between the people reading this book and those who can do yoga moves like the one on the left is very, very small.

Congratulations if you one of the leet few!

Matrix Factorization

For matrix factorization we don't tell the algorithm a preset list of features (female vocal, country, PBR&B, etc.). Instead we give the algorithm a chart (matrix) like the following

and ask the algorithm to extract a set of features from this data.

Customer	Taylor Swift	Miranda Lambert	Carrie Underwood	Jhené Aiko	Kelela
Sarah	5	?	5	2	2
Miguel	2	?	?	4	5
Tyler	5	5	5	2	?
Ann	2	3	?	5	5
Jessica	2	1	?	5	?

To anthropomorphize this yet even more, it is like asking the algorithm,

Okay algorithm, given 2 features (or some number of features) call them feature 1 and feature 2, can you come up with the P and Q matrices?

I comply

Note: I deleted one artist from the table: Tinashe. I am doing this just to reduce the math computations we will be doing by hand on the following pages. I did not have a falling out with Tinashe.

These extracted features are not going to be something like 'female vocals' or 'country influence'. In fact, we don't care what these features represent. Again, we are going to ask the algorithm to extract features that are hidden in that table above. In order to make this sound a bit fancier than 'hidden features' data scientists use the Latin word for 'lie hidden', *lateo*, and call these **latent features**.

Let's get a rough idea of how this works by looking at ratings of musical artists we know nothing about--the famous musical artists *a*, *b*, *c*, *d*, and *e* and we would like to explain these ratings by means of two features which we will arbitrarily call *feature 1* and *feature 2*.

Customer	a	b	c	d	e
Sarah	5		5	2	2
Miguel	2		1	4	5
Tyler	5	5	5	2	1
Ann	2	3		5	5
Jessica	1	2		5	

R

We might reason as follows. It seems that Sarah and Tyler like artists *a*, *b*, and *c* and don't like artists *d*, and *e*. And Miguel, Ann, and Jessica are the exact opposite, liking artists *d* and *e* and not liking *a*, *b*, and *c*. So artists *a*, *b*, and *c* group together. They seem to have something in common so let's label that commonality *feature 1*. And *d* and *e* group together so we might say they share *feature 2*. And we can make finer distinctions. Since Ann and Jessica like artist *b* a bit more than *a* perhaps *b* has more of *feature 2*. This is roughly what the algorithm does.

The inputs to the matrix factorization algorithm are the data in the chart shown above and the number of latent features to use (for example, 2). Our eventual goal is to calculate

$$\hat{R}$$

which is a table of estimated ratings. That is, a table similar to the above but with all the numbers filled in:

Customer	Taylor Swift	Miranda Lambert	Carrie Underwood	Jhené Aiko	Kelela
Sarah	5	**4.78**	5	2	2
Miguel	2	**2.97**	**1.64**	4	5
Tyler	5	5	5	2	**2.52**
Ann	2	3	**1.45**	5	5

Customer	Taylor Swift	Miranda Lambert	Carrie Underwood	Jhené Aiko	Kelela
Jessica	2	1	**1.16**	5	**5.22**

The bolded numbers are those that were blank in the original chart but predicted by our algorithm. The unbolded numbers are predicted values that have an actual value in the original table. From our original data we see that Jake gave a rating of 5 to both Taylor Swift and Carrie Underwood and we see that the algorithm's estimates for those are 4.92 and 4.94 —pretty good! To get these predicted values we use latent features as an intermediary. Let's say we have two features: feature 1 and feature 2. And, to keep things simple, let's just look at how to get Jake's rating of Taylor Swift. Jake's rating is based solely on these two features and for Jake, these features are not equal in importance but are weighed differently. For example, Jake might weigh these features:

The inputs to the matrix factorization algorithm are the data in the chart shown above and the number of latent features to use (for example, 2). Our eventual goal is to calculate

$\hat{R}$

a table of estimated ratings. That is, a table similar to the above but with all the numbers filled in: