**A Proven Method to Remember Data Science Concepts For as Long as You Need**

And tools to put the method into practice in the age of AI

**The problem with self-learning data science**

Every time I want to install a library with Anaconda, the -c part of the command keeps moving around. So, like most people, I google it, sometimes 3-4 times a day:

conda install -c conda-forge library\_name

**A Proven Method to Remember Data Science Concepts For as Long as You Need**

And tools to put the method into practice in the age of AI

**The problem with self-learning data science**

Every time I want to install a library with Anaconda, the -c part of the command keeps moving around. So, like most people, I google it, sometimes 3-4 times a day:

As self-taught data scientists, we have none of that pressure. All we have is our self-discipline that keeps persuading us we are doing an excellent job as we watch a YouTube course on our couch.

Our learning processes are haphazard. We learn something new and jump to the next shiny thing without **the first thing quite penetrating our brains**.

We leave information retention up to chance.

When we actually sit down to practice what we “learned” (air quotes), we’ll realize we already forgot 80% of the new knowledge in the time it took to turn on our computers.

So, we start googling. And after this behavior becomes the norm, we brag to others how we are exceptional at googling in our little tweets. What we are *actually* doing is subtly signaling to others that we have no reliable systems whatsoever to learn and retain the overwhelming amount of information in data science.

Through no fault of our own, we became **the worst type of learners**.

**The solution**

Without effective methods and tools to learn and retain new knowledge, it is tough to become a data scientist.

There is just so much to learn: math, statistics, machine learning theory, the functions and methods in dozens of Python libraries, and so on. It is hard to keep track of all this information.

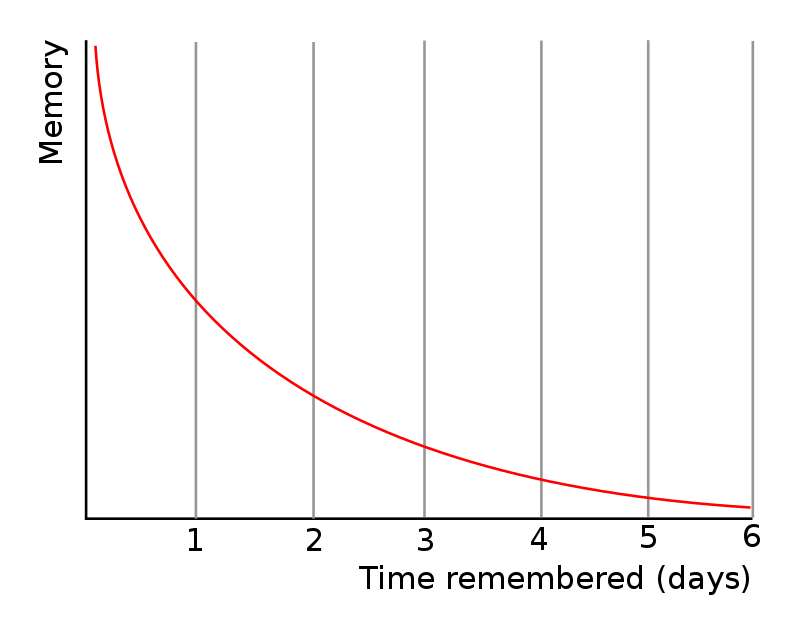


Image by [Wikipedia](https://en.wikipedia.org/wiki/Forgetting_curve#/media/File:Forgetting_curve_decline.svg). Wikimedia commons.

The Ebbinghaus forgetting curve above shows the rate at which new information leaks from memory.

It is clear from the graph that it will take only six days to lose new info completely. And when it is information learned in our haphazard and careless ways, it will become even shorter.

But once you make a serious effort to put new knowledge into a reliable *repetition system*, you consciously choose to remember it for the rest of your life or as long as you need it.

Can I quite possibly be talking about rote learning (🤒)? No, of course not. I am talking about **spaced repetition**!

Spaced repetition is a powerful memory technique that greatly takes advantage of the Ebbinghaus forgetting curve:

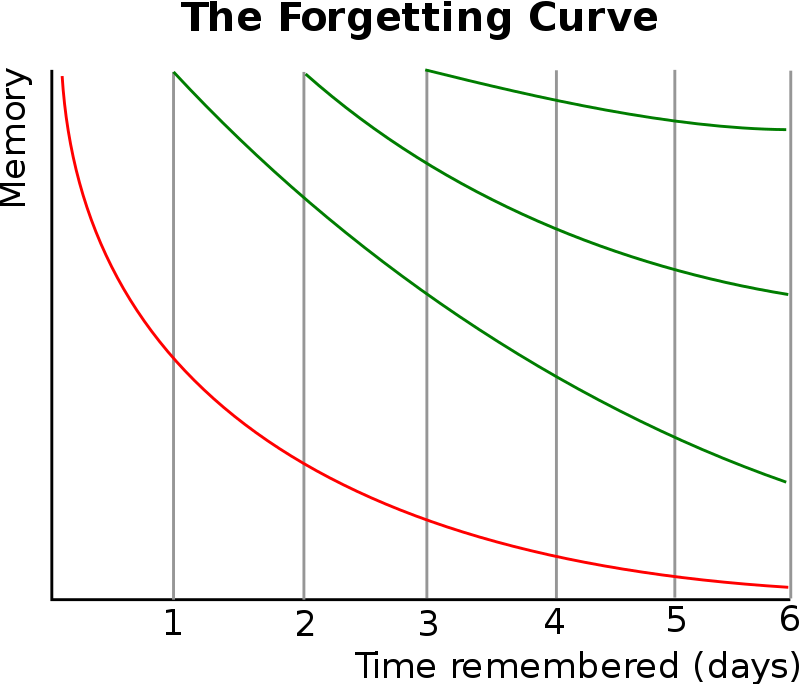


Image by [Wikipedia](https://en.wikipedia.org/wiki/Forgetting_curve#/media/File:ForgettingCurve.svg). Wikimedia commons.

Spaced repetition re-exposes you to new information at increasingly larger optimal intervals, each interval coming just when a memory leak is about to happen.

This will reset your memory and increase the next interval where you have to review the material.

**What are the benefits of SR?**

Perhaps, the most beneficial thing about spaced repetition is the way it transfers knowledge from short to long-term memory.

Apart from the efficient use of time and improved retention, studies show the following benefits of the system:

* Personalization: Customizable to your unique preferences, as it adapts to your pace and level of mastery of the material.
* Improved comprehension: By reinforcing concepts and connections continually over time, it becomes easier for you to build a network of knowledge and understand complex topics more deeply.
* Increased motivation: Spaced repetition gives me a great sense of progress and achievement as my repetition intervals get longer.

These are probably why many medical students swear their lives on this method because they use it to memorize the names of bones, blood vessels, nerve branches, and all the exhausting details about the human body.

Data science may not be as complicated, but we still have a fairly large amount of things to remember.

**Spaced repetition algorithms**

There are many algorithms implementing spaced repetition in practice, the most popular of which is **SuperMemo**.

SuperMemo is a series of SR algorithms that has steadily been coming out since 1982. The author, Dr. Piotr Wozniak, was recognized by Wired magazine as the [“inventor of a technique to turn people into geniuses”](https://www.wired.com/2008/04/ff-wozniak/) in 2008.

So, how do *you* turn into a genius with this method?

After sufficiently learning the underlying concepts and facts, you first break down the material into chunks using flashcards (yes, I understand this is a big problem but **bear with me till the end**).

After making a database of cards, you start to review them in sessions. The first session shows the cards in the order they were added or shuffled (based on your preferences). Then, you rate the cards on how well you recall them.

In SuperMemo-2, ther are six options:

* 0: I have no clue whatsoever
* 1: Incorrect, but after seeing the answer, it rings a bell
* 2: Incorrect, but after seeing the answer, it came rushing back to me
* 3: Correct response, but I had to dig deep and make an effort to remember
* 4: Correct response, but I am hesitating
* 5: I remember it as though it was minutes ago

Then, the chosen rating is plugged into long calculations that involve the number of times the card was successfully recalled before, the easiness factor of the card (don’t ask), and the inter-repetition interval. The final result will determine when the card must be shown again.

For cards rated below 4, SuperMemo will ask you to review the card as many times as you need during the current session until the rating goes above 4.

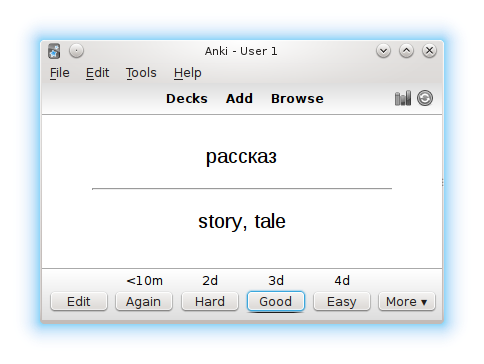
Each correctly recalled card will be shown after increasingly long intervals. For example, if you memorize that the function to convert a timestamp into a datetime is datatime.datetime.fromtimestamp, you only have to review the card showing this information 4–5 times over the span of a month to remember it for the coming six months.

As you might imagine, this is a much better repetition system than rote learning, fixed interval repetition, or worst, repetition when the mood strikes you.

**Spaced repetition tools**

There are many SR tools powered by SuperMemo-like algorithms.

The first (and this one is the king) is [Anki](https://apps.ankiweb.net/" \t "_blank). It is open-source and implements a modified version of SuperMemo-2. Instead of providing six recall ratings, it shows four:



Anki being used to memorize Russian vocab. [Image by Wikipedia](https://en.wikipedia.org/wiki/Spaced_repetition" \l "/media/File:Anki_2.0.22_KDE4.en.story-ru-en.smallwindow.png" \t "_blank). Wikimedia commons.

As it is open-source, it has a very antique look, but it is a cross-platform, free application (except for the iOS version). The [GitHub repo of the software](https://github.com/ankitects/anki) has over 13k stars, which suggests massive support from the community.

They have been working on Anki for over ten years, and the current version has the following features:

1. Available everywhere: Windows, macOS, Linux, Android, and iOS (this one costs money)
2. Fully customizable: create your own flashcards, organize them into decks, and set your own parameters to the spaced repetition algorithm
3. Sync across devices: the computer version of Anki is the main app and mobile and web versions are only companions but synced.
4. Multimedia support: Add images, audio, video, text formatting, and LaTeX to make flashcards memorable and engaging. There is also support for [image occlusions](https://ankiweb.net/shared/info/1374772155" \t "_blank) to memorize visual information.
5. Add-ons: similar to Python extensions, you can create and add your own functionality to the software, like custom keyboard shortcuts, themes, and advanced statistics.
6. Pre-built decks: community constantly shares decks with pre-made cards for popular topics. This includes hundreds of thousands of cards on language learning or virtually any subject in university exams and many other great/cool/weird topics.

One obvious pain point we **didn’t stress** is creating flashcards unavailable in the community.

I know that data science is a relatively young field when it comes to spaced repetition. Anyone would have an enormous amount of information to convert into flashcards, which sounds tedious and sickening. But it is a necessary evil.

I firmly believe that the overall time it takes for you to create flashcards for one topic and **totally master it** with spaced repetition will be much less than hours of googling or dozens of **vicious cycles** of forgetting and relearning.

Besides, we are lucky to be living in the golden age of AI (we are, aren’t we?). There are already cheap AI-powered flashcard software like [Monic.ai](https://beta.monic.ai/).

I already tried Monic.ai, and it looks great. You upload a screenshot or a PDF file, and it automatically converts the text inside into flashcards in mere seconds. It is powered by spaced repetition as well.

If you decide to give it a go, you should consider downloading the [GoFullPage Chrome extension](https://chrome.google.com/webstore/detail/gofullpage-full-page-scre/fdpohaocaechififmbbbbbknoalclacl?hl=en" \t "_blank) to take full-page screenshots or know [how to save web pages as PDFs](https://www.consumerfinance.gov/consumer-tools/save-as-pdf-instructions/" \t "_blank) so that you can turn any online article, tutorial, or documentation page of Python frameworks into flashcards with Monic.ai.

**Wrap**

It is time to change our approaches to learning data science. We should ditch our careless, haphazard ways of watching YouTube videos just for the sake of watching or taking courses back-to-back in search of a new worthless e-certificate.

We should stop learning something once and hope for the best that it stays there. We should stop wishful thinking.

We should stop leaving memory up to chance.

Instead, we should take deliberate actions to memorize every necessary fact, piece of theory, concept, terminal command, Python function, or function argument for as long as we need them.

Yes, this will take some getting used to, but once we are, we can significantly shorten the time it takes to go from “learning data science online” to “doing data science in a job that pays six figures”.

Thank you for reading!

Loved this article and, let’s face it, its bizarre writing style? Imagine having access to dozens more just like it, all written by a brilliant, charming, witty author (that’s me, by the way :).

For only 4.99$ membership, you will get access to not just my stories, but a treasure trove of knowledge from the best and brightest minds on Medium. And if you use [my referral link](https://ibexorigin.medium.com/membership), you will earn my **supernova of gratitude** and a virtual high-five for supporting my work.