In our example, a user may have two states:

Has Time Management (TM)

No Time Management (TM)

These are the hidden states, since we do not have a way to determine time management without some form of event happening. The event we choose to look at:

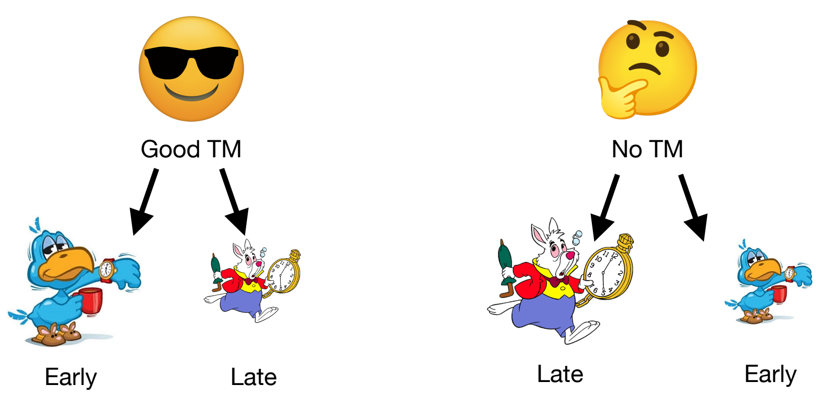
Started lesson early

Started lesson late

We can determine what “early” or “late” is relatively to a deadline or a lesson or the amount of days lapsed from the beginning of a lesson. For example, if there is a week to complete a lesson, we may determine that any first log-in after 20% of the overall time is considered “late”. Or, for a lesson with no deadline we may determine that if a user took more than 3 days to first log-in after they were told of a lesson, then they are “late”. The definition of “early” or “late” may be specific to a lesson or context but for our model it must ultimately be defined somehow.

In developing a model we need to identify the hidden and observable states.

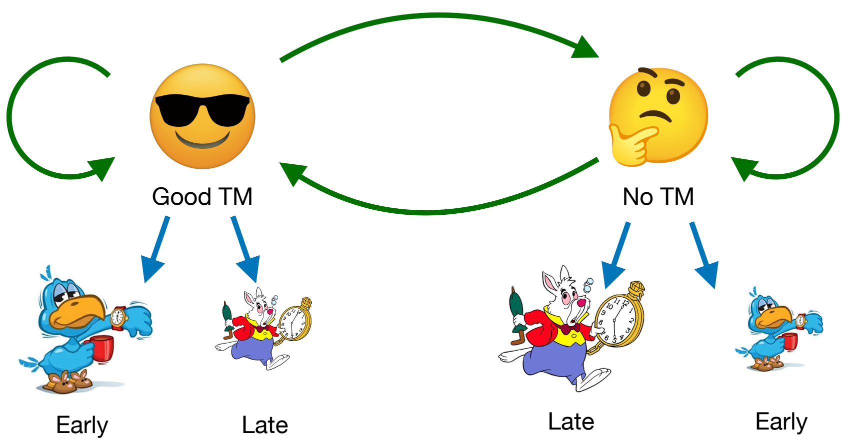
If there was a direct, linear correlation between being early and having time management then no model would be needed. An observation that a user is “early” would lead to the conclusion that they have good time management skills. However in reality even those who have good time management skills are sometimes late, and those who have no time management skills are sometimes early.



In addition to the fact that learners with good time management are sometimes late, time management itself is not a constant state. For example, one could have good time management skills when they are relaxed, but when they are stressed they become distracted and unable to manage their time. Or, one could have better ability to self regulate, and therefore manage their time, when they are interested in a subject. When the subject is not interesting to them they may become unmotivated and unable to manage their time. The link between time management and motivation, for example, can and should be examined later. For now, we must only determine if the learner has good time management skills or not.

To accomplish this we must add to our model the following elements:

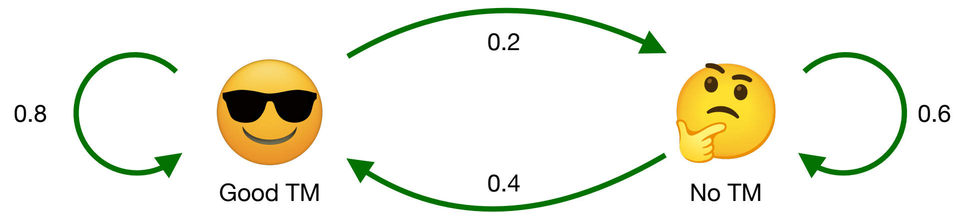
(1) Transition probabilities (green arrows). These describe the probability of shifting between having good time management skills to having bad TM skills, and vice versa.
(2) Emission probabilities (blue arrows). These describe the probability of a state based on the hidden state. For example, the probability of a learner with good time management skills to be early. Since these probabilities are a result of the hidden states, they are "emitted" from them.



<u>Transition probabilities</u>
These probabilities describe the chances of "this occurrence" being different from "last time". In our case, the probability of a learner who has good time management skills to still have them this time as well. There are a few ways to arrive at the answer. Over time, we will probably see the types of contexts affecting learners with time management skills. For example, if the lesson is on the same subject that a learner exhibited good time management skills + have good motivation then it is more likely that they will still have good time management skills. In the future, the output of other models (motivation, for example) could have an effect on the transition probability of this model.

For now, we can simply look at the data collected to find learners who exhibit good time management skills. But we are looking at the "transitions" this time. In other words, how many times did a "Good TM" transition to a "No TM", and how many times did a "Good TM" stay in the same state, across lessons. For example, if out of 6 lessons (5 total transitions) a "Good TM" learner transitioned to "no TM" one time, there is a 20% transition probability from "Good TM" to "No TM" and a 80% transition probability from "Good TM" to "Good TM" again. The same process can be repeated for those with no good time management skills. Over time as more data arrives, these probabilities can be refined and improved.

If no data exists then a preliminary probability can be used. For example, let's say that there is an 80%, or 0.8 probability that those with good time management skills will stay in the same state and a 0.2 probability that they will transition to the different state of "No TM." And also that the similar, but not identical, transition probability exists for those with "No TM" (although, according to research self-regulation can be learned and therefore time management can be improved. Here, also, the probability can be a result of a more complex calculation which takes into account the time a user spent on the platform and being allowed to grow and improve their time management skills).



Emission probabilities

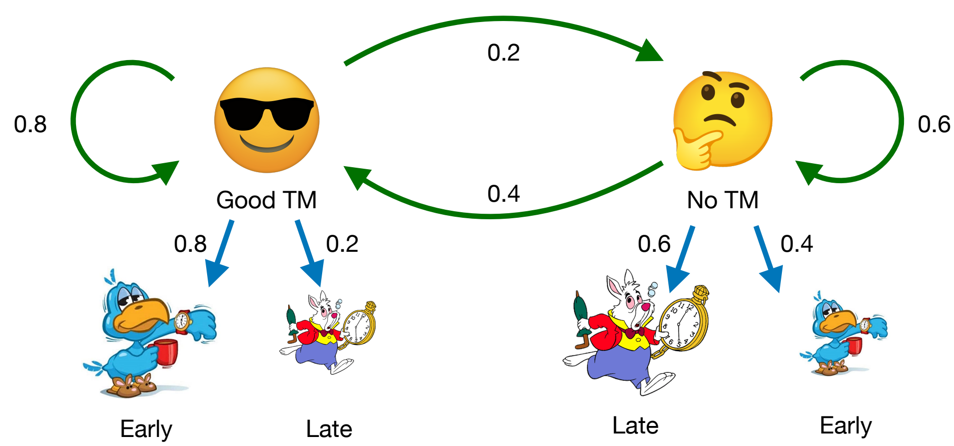
Similarly to the transition probabilities, the emission probabilities can be learned from previous data. The questions we ask are:

(1) How many times was a "Good TM" learner late? How many times were the early?
(2) How many times was a "No TM" learner late? How many times were the early?

The answers to these questions feed the model. For example, if a "Good TM" learner was early 4 times and late 2 times, then there is a 67%, or 0.67, probability that a "Good TM" learner will be early and a 0.33 probability that they will be late.

Again, as more data is generated the model is improved. In addition, additional circumstances can affect these probabilities. We acknowledged that lack of motivation, for example, can cause a transition from "Good TM" to "No TM" for some people. But are there situations where a learner with good time management skills can still be late? Identifying these and factoring them in the model can increase its complexity and make it more accurate.

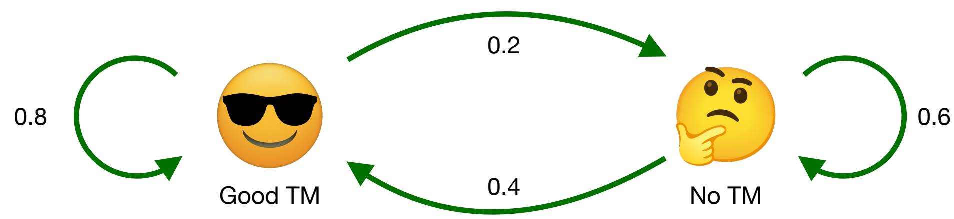
For now, let's assume some preliminary values based on experience with students. This is our model with the probabilities included:



Using the model
When developing models we should ask ourselves when the model can be called. There can be two situations where this model is used:

(1) Making a decision without any event (observation) from the learner. For example, we are dynamically building a lesson for a user and they haven't logged-in yet. Even though they are not yet there, we need to make certain decisions predicting their time management abilities based on the data we have.
(2) Making a decision with an event (observation) from the learner. For example, the learner was early. But we need to know what that means and the model will provide that answer.

No observation
The question we are trying to answer is whether the learner is "Good TM" or "No TM". In the absence of an observation, we are left with the general probability that is embedded in our model. Specifically it is the transition probability we are concerned with.



Even though it may be the learner's first time accessing the platform, they do have an established state of time management and we can imagine that this is not the first time they need to use this skill. Therefore we can look at the probability of a learner being "Good TM" by adding the probability of "Good TM" repeating itself (0.8) or followed by a "No TM" (0.3).

User likely to be **Good TM = (Good TM) 0.8 + (No TM) 0.4**

Similarly we can calculate the probability of "No TM" by adding up the probabilities leading to "No TM" in our model:

User likely to be **No TM = (Good TM) 0.2 + (No TM) 0.6**

Since the user can either be fully in one of the two states, "Good TM" or "No TM", one will always be equal to 100% when the other is 0%. Or:

(Good TM) + (No TM) = 1

Using only one of the "likely" equations above should be enough to solve the probability of a user being either "Good TM" or "No TM". In this case:

No TM = 0.33
Good TM = 0.67

Based on this probability we can make decisions for the learner without any observation. As the data that feeds the transition probability improves, so will our prediction.

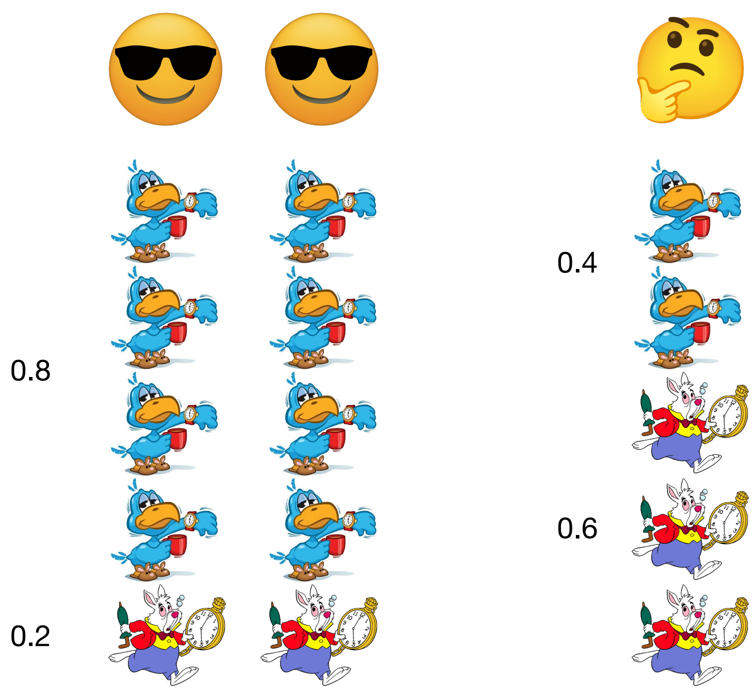
Observed event

In the case of an observed event (the user logged-in, either early or late) we can use the model to infer meaning. If the learner was early, it does not necessarily mean that the learner is of type “Good TM” because even “No TM” learners are sometimes early. By calculating the meaning of the event we are able to give the learner a score relevant to time management (increase or decrease it) and make decisions based on that probability.

Since the event in question is singular (we are only looking at this one even, not a sequence of events), we care less about the transitional probabilities in our model. However, the transitional probabilities were useful to us in calculating the likelihood of a random leaner (one with no history in our platform) being either “Good TM” or “No TM”. The calculation used for the observed events needs these calculated parameters as *initial probabilities* for the equation (the fact that the learner can be “Good TM” with a probability of 0.67 or “No TM” with a probability of 0.33, in our case). The calculation itself is part of Bayes' theorem which is concerned with the probability of an event, based on prior knowledge of conditions that might be related to the event.

According to the initial properties, the ratio between “Good TM” and “No TM” is 2:1.

When a learner is in a “Good TM” state they can be either early or late. According to our model, there is a probability of 0.8 that they will be early and 0.2 they will be late. Similarly, for “No TM” there is a probability of 0.4 that they will be early and 0.6 they will be late.



The purpose of this illustration is to connect between the two different domains in which a learner can be “early”. If a learner is “early” then they are either in the group belonging to

“Good TM” (a total of 8) or in the group belonging to “No TM” (a total of 2). Adding the two groups together equals 10. Therefore the probability of a learner who is “early” to belong to the “Good TM” group is 8/10 or 0.8. The probability of them belonging to the “No TM” group is 2/10, or 0.2.

If the learner is “late” then they can either belong to the “Good TM” group (a total of 2) or the “No TM” group (a total of 3). Therefore the probability of a learner who is “late” to be of a “No TM” state is 3/5 or 0.6.

These probabilities are called the posterior probabilities in Bayesian statistics as they are updated from the initial probabilities when new information (the event) is taken into account.

If we used the “no observation” method to predict whether a new user is of “Good TM” we would arrive at 0.67 since this is based only on our assumptions. Once the learner entered the system, if the are “early” then we update our score to 0.8 (based on the collation above) which is the probability of them being of “Good TM”.

However this method only takes into account one previous state. Once there is a longer series of events, a different calculation should be used.

Two event series

We examined the use of the model when no event (observation) occurs and when there is one observed event. If there is more than one observed event then the probability should be different due to the emerged pattern, even if short.

Let’s examine a situation in which there are two lessons and two observations:

Lesson A

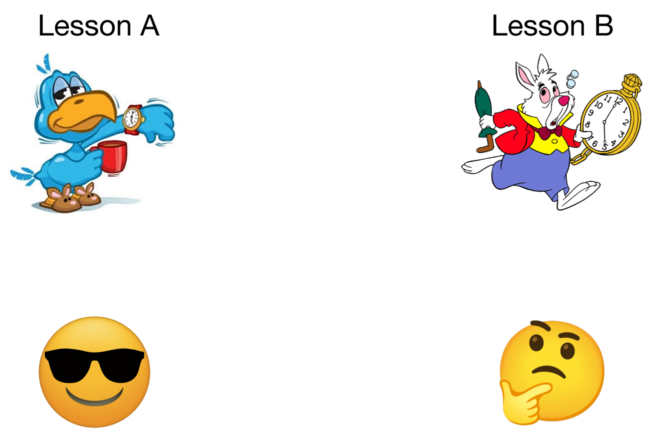
Lesson B

The learner was early in the first lesson and late in the second lesson. It is up to the model to calculate whether there is a higher probability of the learner to be “Good TM” or “No TM” based on this information.

Ignoring all other factors (like those which may increase the likelihood of a transition between “Good TM” to “No TM” like reduced motivation) there can be four different possibilities in the transition between the two lessons:

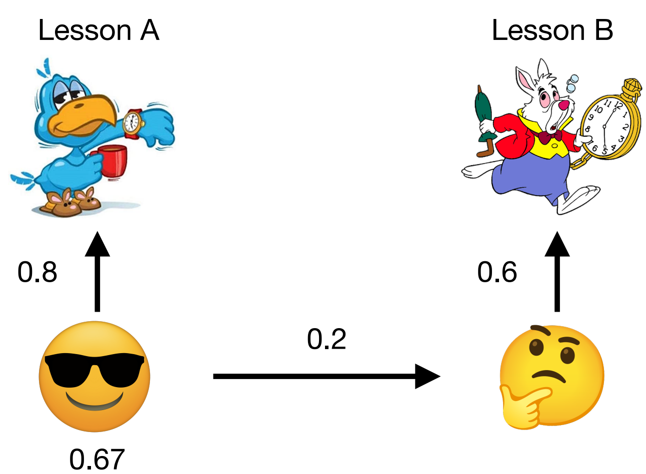
The above illustration covers all possibilities in the transition of the hidden states between the two lessons. The calculation to be performed, called *Maximum Likelihood Estimation* simply calculates the probability for each one of the transitions given the observations. The one with the highest probability is the one chosen as output.

An example:



This example assumes one of the four transition possibilities: The learner had “Good TM” in lesson A and transitioned to “No TM” in lesson B. To calculate the probability of this specific transition happening, we need to recall the parameters calculated up to this point:

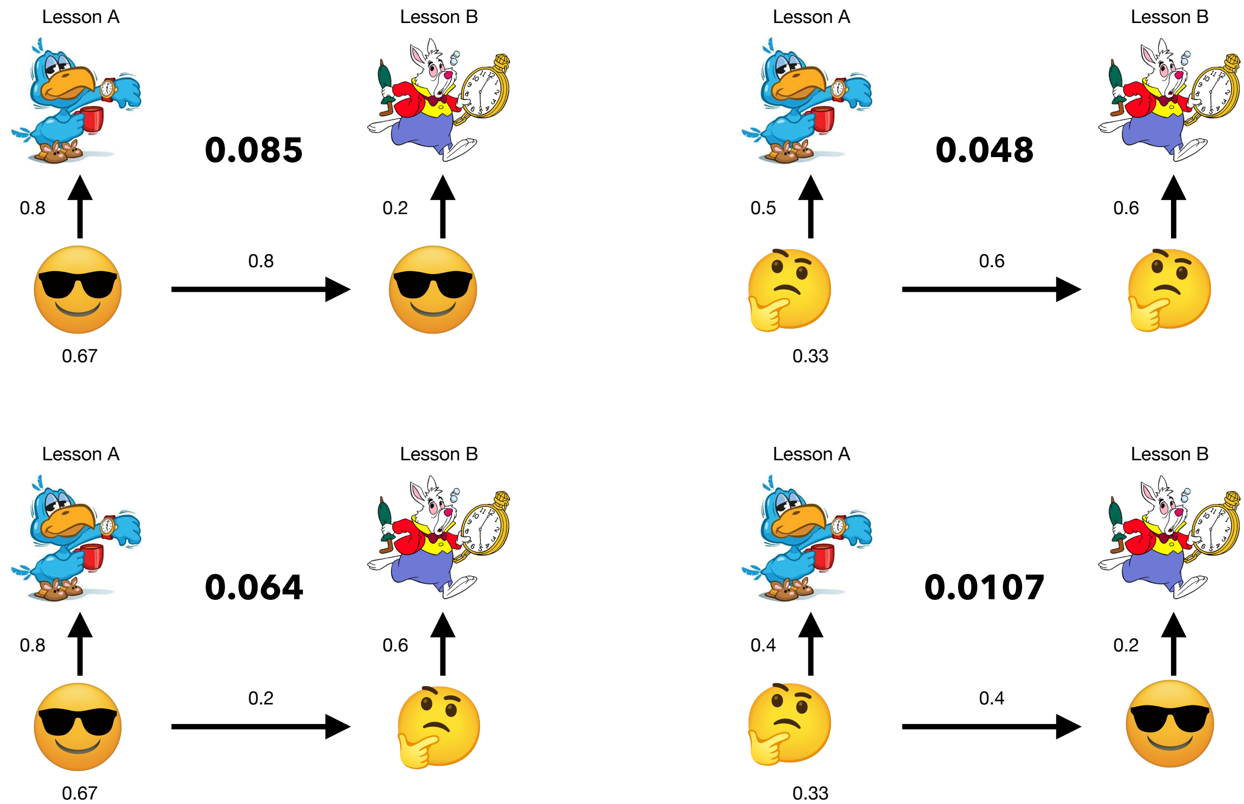
(1) The probability that the learner is “Good TM” without any prior observation is 0.67.
(2) The probability of “Good TM” being early in lesson A is 0.8.
(3) The transition probability going from “Good TM” to “No TM” is 0.2.
(4) The probability of “No TM” being late is 0.6.



The probability of the entire situation occurring is the product of multiplying all the probabilities together.

0.67 x 0.8 x 0.2 x 0.6 = <u>0.064</u>

Maximum Likelihood Estimation demands that we compare all four scenarios and pick the highest probability:



Based on these calculations, the model that is most likely to happen is the highest score, or probability of 0.085 in which the learner maintains the state of "Good TM" even though they were late in Lesson B. However, since we are interested in updating a score we can subtract the portion of 0.085 from the first score of 0.8 (the initial score for being early) (?):

$$\mathbf{0.8 - 0.085 = 0.715}$$

While the learner is still in the realm of "Good TM", their score was affected by being late.

<u>Multiple event series</u>
If the event series is more than one, we must be able to mathematically calculate the probability. To implement this in code, we need to represent the data we have so far (transition probabilities and emission probabilities) in the forms of tables, or matrixes:

Transition Matrix

	😎	🤔
😎	0.8	0.2
🤔	0.4	0.6

Emission Matrix

😎	0.8	0.2
🤔	0.6	0.4