

learning rate

[Document subtitle]



April 1, 2020

akram ayman

2015978004

In [machine learning](https://en.wikipedia.org/wiki/Machine_learning) and [statistics](https://en.wikipedia.org/wiki/Statistics), the learning rate is a [tuning parameter](https://en.wikipedia.org/wiki/Hyperparameter_(machine_learning)) in an [optimization algorithm](https://en.wikipedia.org/wiki/Mathematical_optimization) that determines the step size at each iteration while moving toward a minimum of a [loss function](https://en.wikipedia.org/wiki/Loss_function).[[1]](https://en.wikipedia.org/wiki/Learning_rate#cite_note-1) Since it influences to what extent newly acquired information overrides old information, it metaphorically represents the speed at which a machine learning model "learns."[[2]](https://en.wikipedia.org/wiki/Learning_rate#cite_note-towardsdatascience2-2) The learning rate is often denoted by the character η or α.

In setting a learning rate, there is a trade-off between the rate of convergence and overshooting. While the direction toward the minimum is usually determined from the [gradient](https://en.wikipedia.org/wiki/Gradient_descent) of the loss function, the learning rate determines how big a step is taken in that direction.[[3]](https://en.wikipedia.org/wiki/Learning_rate#cite_note-3) A too high learning rate will make the learning jump over minima but a too low learning rate will either take too long to converge or get stuck in an undesirable local minimum.

In order to achieve faster convergence, prevent oscillations and getting stuck in undesirable local minima the learning rate is often varied during training either in accordance to a learning rate schedule or by using an adaptive learning rate.[[4]](https://en.wikipedia.org/wiki/Learning_rate#cite_note-variablelearningrate-4) In [Newton's method](https://en.wikipedia.org/wiki/Newton%27s_method_in_optimization), the learning rate is essentially determined from the local curvature of the loss function, by using the [inverse](https://en.wikipedia.org/wiki/Invertible_matrix) of the [Hessian matrix](https://en.wikipedia.org/wiki/Hessian_matrix) as the step size.

A learning rate schedule changes the learning rate during learning and is most often changed between epochs/iterations. This is mainly done with two parameters: decay and momentum . There are many different learning rate schedules but the most common are time-based, step-based and exponential.[[4]](https://en.wikipedia.org/wiki/Learning_rate#cite_note-variablelearningrate-4)

Decay serves to settle the learning in a nice place and avoid oscillations, a situation that may arise when a too high constant learning rate makes the learning jump back and forth over a minima, and is controlled by a hyperparameter.

Momentum is analogous to a ball rolling down a hill; we want the ball to settle at the lowest point of the hill (corresponding to the lowest error). Momentum both speeds up the learning (increasing the learning rate) when the error cost gradient is heading in the same direction for a long time and also avoids local minima by 'rolling over' small bumps. Momentum is controlled by a hyper parameter analogous to a ball's mass which must be chosen manually—too high and the ball will roll over minima which we wish to find, too low and it will not fulfil its purpose. [The formula for factoring in the momentum](https://en.wikipedia.org/wiki/Stochastic_gradient_descent#Momentum) is more complex than for decay but is most often built in with deep learning libraries such as [Keras](https://en.wikipedia.org/wiki/Keras" \o "Keras).

Time-based learning schedules alter the learning rate depending on the learning rate of the previous time iteration. Factoring in the decay the mathematical formula for the learning rate is.

Step-based learning schedules changes the learning rate according to some pre defined steps. The decay application formula is here defined as.

Exponential learning schedules are similar to step-based but instead of steps a decreasing exponential function is used. The mathematical formula for factoring in the decay is.

معدل التعلم

معدل التعلم: هو مقياس شائع يتحكم في مقدار تغيير النموذج استجابة للخطأ المقدر في كل مرة يتم فيها تحديث أوزان النموذج.

وهو يقوم بتحديد حجم الخطوة في كل تكرار اثناء التحرك نحو الحد الادنى , وهو يمثل مقياسا عاما لسرعة التي يتعلم بها النموذج وغالبا ما يتم الاشاره اليه بالحرف

η or α.

في تحديد معدل التعلم, هنالك مفاضلة قائمة بين معدل التقارب وتجاوز الحد من خلال حجم الخطوة المختارة بناءا على قيمة معدل التعلم,فأن قيمة معدل التعلم المرتفعة تجعل الخطوة كبيرة مما يشكل خطر في تجاوز القيمة المطلوبة المحققة للهدف وكذلك القيمة المنخفضة ستسغرق وقتا طويلا للالتقاء بالقيمة المطلوبة او التعثر بقيمة صغيره غير مرغوب فيه. من أجل تحقيق نتيجة أسرع, ومنع التذبذبات الخاصه في معدل التعلم غالبا ما يتغير معدل التعلم اثناء تدريب إما وفقا لجدول معدل التعلم أو باستخدام معدل التعلم التكيفي.وفي طريقة نيوتن, يتم تحديد معدل التعلم بشكل اساسي من الانحناء المحلي لوظيفة الخسارة, وذلك باستخدام معكوس المصوصفة لحجم الخطوة.

جدول معدل التعلم:

يغير جدول معدل التعلم أثناء التعلم وغالبا ما يتم تغييره بين التكرارات.ويتم ذلك اساسا مع اثنين من المعاملات:وهما الزخم والاضمحلال, وهنالك العديد من الجداول الزمنية لمعدلات التعلم المختلفة, لكن الاكثر شيوعا هي الجداول الزمنية والأسية وجداول خط الاساس.

الزخم:هو ايجاد تسوية القيمة عند اقل نقطة لأقل خطأ, يعمل الزخم على تسريع عملية التعلم عندما يتدرج تدرج تكلفة الخطأ في نفس الاتجاه لفترة طويلة وتجنب الحد الادنى المحلي عن طريق تجاوز العثرات الصغيره. يتم التحكم في الزخم بواسطة المعاملات المتطورة او العليا ال (hyperparameter)

ويتم اختيارها قيمة عالية في البداية وبعد ذلك نتيجة التكرار وتغير في مقدار معدل التعلم تبداء بالانخفاض الى ان تحقق الغرض

الاضمحلال : يعمل الاضمحلال على تسوية التعلم وتجنب التذبذبات الناتجة, وهو الموقف الناتج عن ثبوت معدل التعلم ويتم التحكم به بواسطة المعاملات المتطورة ال(hyperparameter)

الجداول الزمنية : تعمل هذه الجداول على تغيير معدل التعلم اعتمادا على معدل التعلم في التكرار الزمني السابق اعتمادا على العلاقة الرياضية التاليه:



:حيث أن

N= معدل التعلم

d= معامل الاضمحلال

n= عدد الخطوات

جداول الخطوة : تتغير جداول التعلم معدل التعلم وفقًا لبعض الخطوات المحددة مسبقًا. يتم تعريف صيغة تطبيق الاضمحلال على النحو التالي :



حيث أن :

n= معدل التعلم

n.= معدل التعلم الابتدائي

d= مقدار التعلم المتغير مع كل تكرار

r= معدل الانخفاض

تتشابه جداول التعلم الأسي مع الخطوات المستندة إلى الخطوة ، ولكن بدلاً من الخطوات ، يتم استخدام دالة الأس المتناقص. الصيغة الرياضية للعوملة في الاضمحلال هي :



New link : <https://ar.wikipedia.org/wiki/%D9%85%D8%B9%D8%AF%D9%84_%D8%A7%D9%84%D8%AA%D8%B9%D9%84%D9%85_(learning_rate)>