

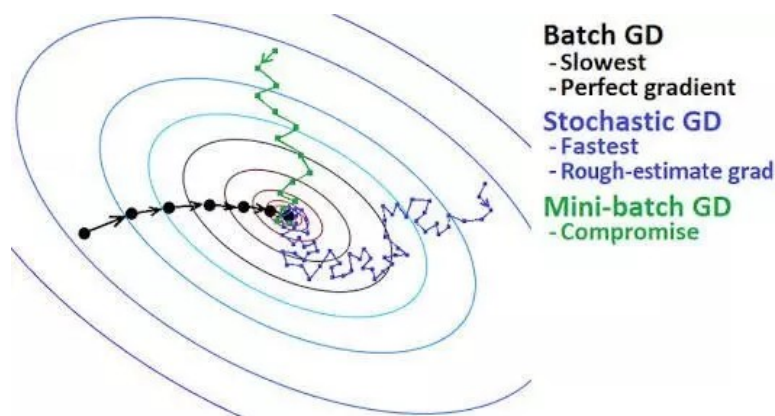
انواع گرادیان کاهشی

Batch Gradient Descent

برای محاسبه GD معمولی از تابع هزینه نسبت به یک پارامتر مشتق میگیریم. عبارت دیگر مقدار تغییر تابع هزینه به ازای تغییر آن پارامتر خاص را حساب میکنیم. که به آن مشتق جزئی نسبت به فلان پارامتر (مثلا مشتق y نسبت به x) میگوییم. مثل اینکه پرسیم، اگر من رو به شمال ایستاده باشم، شیب روبرویم چقدر است و اگر رو به شرق ایستاده باشم، شیب پیش رویم چقدر است؟ میتوان به یکباره تمام مشتق های جزئی را محاسبه نمود. وکتور گرادیان برای هر پارامتر مشتق جزئی متناسب با پارامتر (از تابع هزینه) را شامل میشود. دلیل اینکه به آن batch میگویند این است که به یکباره تمام محاسبات انجام میشود. (یعنی به صورت دسته ای) [1]

Stochastic Gradient Descent

SGD یک الگوریتم بهینه سازی محبوب برای مدل های یادگیری عمیق مجموعه داده های بزرگ است زیرا کارآمد است، می تواند به مجموعه داده های بزرگ تر مقیاس بندی شود و با داده های جدید به روزرسانی شود، و در مقایسه با سایر روش های بهینه سازی کمتر در مینیموم های محلی گیر می کند. به دلیل بالا بودن حجم محاسبات در BGD روش موسوم به SGD طراحی شده است. Stochastic به معنای تصادفی است. بنابراین میتوان گفت مزیت SGD سرعت بیشتر و بار محاسباتی کمتر برای یافتن Min است. [1]



Mini-Batch Gradient Descent

روش MBGD چیزی بین دو روش قبلی است. در این روش به جای آموزش کل نمونه ها و یا یک نمونه، هر دفعه یک Batch از داده ها را آموزش میدهم. مثلا اگر کل داده ها N تا است، ۲۰۰ تا ۲۰۰۰ تا آموزش میدهم. در این روش نگرانی از جهت نویزی شدن وجود ندارد. [2]

منبع:

<https://www.linkedin.com/pulse/%DA%AF%D8%B1%D8%A7%D8%AF%DB%8C-%D8%A7%D9%86-%DA%A9%D8%A7%D9%87%D8%B4%DB%8C-shahram-moradpour> .[1]

<https://virgool.io/@danialfarsy/%D8%A8%D8%B1%D8%B1%D8%B3%DB%8C-%D9%88-%D9%85%D9%82%D8%A7%DB%8C%D8%B3%D9%87-batch-gradient-descentmini-batch-gradient-descentstochastic-gradient-descent-n4yklzivliiw> .[2]