

: Perceptron

مقایسه روش Batch و Online :

از نظر حجم محاسباتی : در روش Batch از آنجا که هر iteration روی کل داده ورودی انجام می‌شود، حجم محاسبات در هر تکرار بیشتر از روش Online است که هر مرحله تکرار آن روی یک داده انجام می‌شود. به صورت دقیق با توجه به کد موجود در نوت بوک می‌توان گفت حجم محاسبات برای هر iteration روش Batch و Online به ترتیب $O(N \times d)$ و $O(d)$ است^۱ که N تعداد نمونه‌ها و d تعداد ویژگی‌ها است. حال در همه‌ی تکرارها اگر فرض کنیم برای روش آنلاین روی همه‌ی نمونه‌ها یک تکرار انجام داده و در روش Batch، I تکرار صورت می‌گیرد پیچیدگی زمان روش Batch و Online به ترتیب $O(N \times I \times d)$ و $O(N \times d)$ خواهد بود. از نظر نرخ همگرایی : با توجه به اینکه در روش Batch در هر تکرار همه‌ی نمونه‌ها و تخمینی خوبی از توزیع داده را داریم انتظار داریم که همگرایی به ازای تعداد دفعات تکرار کمتر رخ دهد؛ بنابراین می‌توان گفت سرعت همگرایی روش Batch نسبت به Online بیشتر است ولی در عین حال حجم محاسباتی بیشتری دارد. تحلیل خروجی‌ها : پس از اجرا همانگونه که انتظار داشتیم روش Batch سریع‌تر همگرا شده (در ۴ مرحله برای نرخ 0.1)؛ در نهایت برای نرخ یادگیری 0.1 در روش Batch ۳ خطا برای داده آموزش و ۴۰ خطا برای داده validation داریم که این میزان برای روش Online به ترتیب ۵ و ۴۱ است؛ بنابراین می‌توان گفت از لحاظ صحت با توجه به شرط توقف گفته شده دو روش تفاوت چندانی نداشته‌اند. برای نرخ یادگیری 0.1 مسئله همگرا می‌شود اما برای نرخ 10^{-7} با توجه به کم بودن زیاد آن خطا کم نشده و با توجه به شروط توقف گفته‌شده به جواب بهینه همگرا نمی‌شود.

: Logic Regression

چرایی استفاده از Cross-Entropy به جای MSE : یک دلیل ساده اینست که در صورت استفاده از Cross-Entropy تابع هزینه و مسئله ما در Logistic Regression فرم محدب به خود خواهد گرفت و می‌توانیم با استفاده از روش گرادین کاهشی مسئله را حل کنیم، حال آنکه در صورت استفاده از خطای مربع دو این اتفاق رخ نمی‌دهد. یک دلیل دیگر هم اینست که در مسایل دسته‌بندی باینری بهینه‌سازی لایکی هود لگاریتمی ما را به Cross-Entropy می‌رساند. تحلیل خروجی‌ها : در صورت حذف نکردن داده‌های پرت خطای قابل ملاحظه‌ای مشاهده می‌شود اما با حذف آنها خطا با تقریب خوبی در نهایت به صفر میل می‌کند. نکته دیگر نیز این است که اگر نرخ یادگیری پایین باشد خطا برای تعداد ثابت تکرار بیشتر می‌شود همچنین با افزایش تعداد تکرارها همانگونه که انتظار می‌رود جداسازی بهتر رخ می‌دهد و دقت بالا می‌رود و برای داده تمیز شده (حذف داده‌های پرت) به ۱۰۰ درصد میل می‌کند.

: Linear Regression

مقایسه دو خطای MSE و MAE : خطای MAE نسبت به داده‌های پرت حساسیت بسیار کمتری نسبت به MSE دارد و اگر داده پرت زیاد باشد نسبت به MSE ارجحیت دارد؛ اما اگر داده‌های دور داده‌ی پرت نباشند و آنها در آزمایش ما مورد توجه باشند بهتر است از MSE استفاده کنیم. نکته دیگر اینکه قدر مطلق در صفر مشتق پذیر نیست و این یک مضرت برای MAE محسوب می‌شود. همچنین محاسبات تئوری با MSE راحت‌تر است. تحلیل اجرا : برای نرخ یادگیری بالا مسئله همگرا نخواهد شد اما در نرخ یادگیری کمتر 10^{-6} ، خطا هم برای داده تست و هم آموزش با تکرار در حال کم شدن است و در نهایت به حد معقولی خواهد رسید که نشان از درستی روش دارد. (زمان اجرا

^۱ فرض کرده‌ایم ضرب دو عدد در $O(1)$ انجام می‌شود.

نیز زیاد است.)
نکته : تمام نمودارها در نوت‌بوک موجود است و در این گزارش صرفاً به سوال‌های خواسته شده پاسخ داده شده است.