

به نام خدا

سوال 3: مقدار  $\beta_1$  را میتوان 0 گذاشت تا فقط مقدار  $dx$  لحاظ شود و مومنتوم را اعمال نکنیم اما  $\beta_2$  را اگر 1 بگذاریم که  $dx^2$  حذف شود آنگاه مخرج کسر  $\text{second\_unbias}$  صفر میشود که مجاز نیست:

$$\text{second\_unbias} = \text{second\_moment} / (1 - \beta_2^{**} t)$$

مگر اینکه bias correction را از الگوریتم حذف کنیم و مجاز به تغییر  $\text{second\_moment}$  باشیم آنگاه:

$$\beta_1 = 0$$

$$\beta_2 = 1$$

$$\text{Second\_moment} = 1$$

در این صورت الگوریتم تبدیل به sgd میشود.

البته برای حالتی که bias correction داریم یک ایده در ذهن داشتم که آن را هم بیان میکنم.

$$\beta_1 = 0$$

$$\text{Second\_moment} = 1 - \beta_2^{**} t$$

در این صورت مقدار  $\beta_2$  هر چیزی میتواند باشد ولی در صورتی که مجاز به تغییر  $\text{second\_moment}$  باشیم.

قسمت تحلیل  $\beta_1$  و  $\beta_2$  نوت بوک برای بهینه سازی تابع هدف  $y^2 + x^2$ :

برای  $\beta_1$  و  $\beta_2$  از مقدار 0.9 تا 0.999 را امتحان کردم که 9 حالت را نتیجه داد.

به صورت پیش فرض نرخ یادگیری آدام 0.001 میباشد و در ابتدا با این مقدار نتایج را نمایش دادم که بهترین نتیجه ها برای  $\beta_1 = 0.9$  و  $\beta_2 = 0.9, 0.99$  بود که به مینیمم سراسری 0 رسیدند.

برای نرخ یادگیری با مقدار 0.01 بهترین نتیجه ها برای  $\beta_1 = 0.9$  و  $\beta_2 = 0.9, 0.99, 0.999$  بود که همگی به مینیمم سراسری 0 رسیدند. این مقادیر نشان میدهند که با بزرگتر گرفتن  $\beta_2$  یعنی در نظر گرفتن سهم بیشتر از گذشته شدت تغییرات loss به نسبت شدت تغییرات فعلی loss، نتایج بهتری را میگیریم. در حالیکه در  $\beta_1$  با در نظر گرفتن سهم بیشتر از گرادیان فعلی به نسبت جمع گرادیان های قبلی نتیجه بهتری میگیریم.

#### سوال 4:

از نرخ یادگیری 0.001 تا 0.9 شروع به آموزش و ارزیابی مدل کردیم و بهترین نتیجه برای نرخ یادگیری 0.4 بود که دقت مدل در داده های ولیدیشن حدودا به 98 درصد رسید. بدترین دقت برای نرخ یادگیری 0.001 میباشد که دقت آن رو داده ولیدیشن 82 درصد است و این ینی 16 درصد اختلاف با بهترین دقت. پس انتخاب نرخ یادگیری مناسب روی نتیجه بسیار مهم است.