

پروژه سوم – بررسی شیب

شما بخشی از یک تیم هستید که برای در دسترس قرار دادن پرداخت های تلفن همراه در سطح جهانی کار می کنید و از شما خواسته می شود یک مدل یادگیری عمیق برای کشف تقلب پیاده سازی کنید - هر زمان که شخصی پرداختی را انجام می دهد، می خواهید ببینید که آیا پرداخت جعلی است یا نه.

اما پیاده سازی پس انتشار (Backpropagation) شبکه کاملاً چالش برانگیز است و گاهی اشکالاتی نیز دارد. از آنجا که مأموریت مهم و حساسی در دست دارید، مدیر عامل شرکت شما می خواهد در مورد صحیح بودن اجرای پس انتشار شما اطمینان حاصل کند. مدیر عامل شرکت شما می گوید: "به من اثبات کنید که پس انتشار شما به خوبی کار میکند!" برای دادن این اطمینان ، شما می خواهید از "بررسی شیب" استفاده کنید.

مرحله صفر: اضافه کردن کتابخانه های لازم

محتویات پوشه Requirements را در مسیر پروژه اضافه کنید.

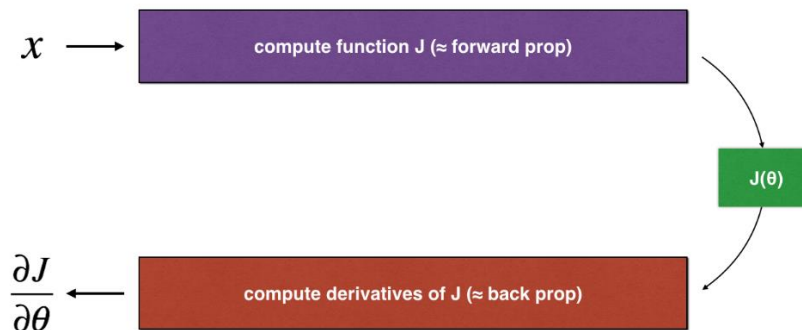
مرحله اول: بررسی شیب چگونه کار میکند؟

پس انتشار ، شیب ها را محاسبه می کند ، جایی که θ پارامترهای مدل را نشان می دهد و J تابع خطای شما را نشان میدهد که در مرحله انتشار رو به جلو محاسبه می شود.

از آنجا که پیاده سازی انتشار رو به جلو به آسانی انجام می شود ، اطمینان دارید که این کار را درست انجام داده اید و بنابراین تقریباً ۱۰۰٪ مطمئن هستید که هزینه J را به درستی محاسبه می کنید. در نتیجه، میتوانید از محاسبه خود برای J کمک بگیرید و محاسبه شیب $\frac{dJ}{d\theta}$ را تایید کنید.

$$\frac{\partial J}{\partial \theta} = \lim_{\epsilon \rightarrow 0} \frac{J(\theta + \epsilon) - J(\theta - \epsilon)}{2\epsilon}$$

تابع خطی یک بعدی $J(\theta) = \theta x$ را در نظر بگیرید، این مدل فقط یک مقدار واقعی (θ) دارد و مقدار x را به عنوان ورودی میگیرد.



نمودار قبل مراحل محاسبه را نشان میدهد. در مرحله انتشار رو به جلو ورودی x دریافت شده و مقدار J محاسبه می شود. در مرحله پس انتشار مشتق $\frac{dJ}{d\theta}$ محاسبه می شود.

تمرین اول: توابع انتشار روبه جلو و پس انتشار را برای تابع یک بعدی گفته شده پیاده سازی کنید.

مرحله دوم: بررسی شیب یک بعدی

برای اینکه بررسی کنیم محاسبه پس انتشار به درستی کار می کند از بررسی شیب استفاده می کنیم. پارامترهای زیر را در نظر بگیرید:

1. $\theta^+ = \theta + \epsilon$
2. $\theta^- = \theta - \epsilon$
3. $J^+ = J(\theta^+)$
4. $J^- = J(\theta^-)$
5. $gradapprox = \frac{J^+ - J^-}{2\epsilon}$

تخمین مشتق از روش زیر به دست می آید:

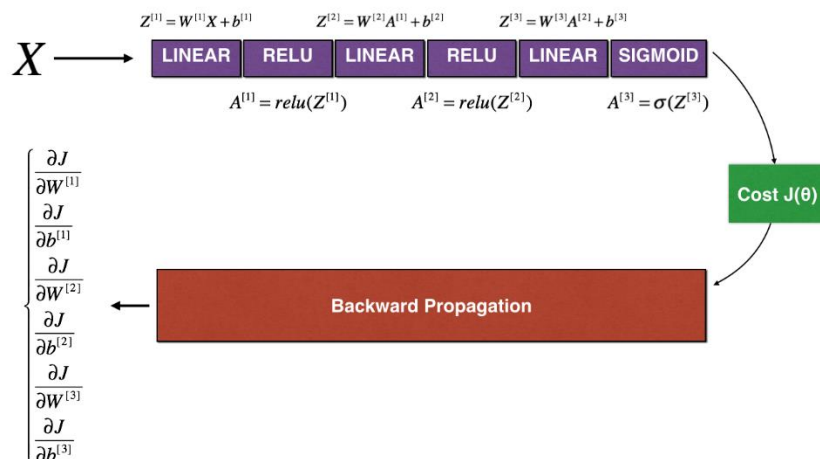
$$difference = \frac{\|grad - gradapprox\|_2}{\|grad\|_2 + \|gradapprox\|_2}$$

کاری که در روش بررسی شیب انجام می‌دهیم مقایسه خروجی پس انتشار با تخمین مشتق است. حد آستانه اختلاف آنها را برابر با 10^{-7} در نظر میگیریم.

تمرین دوم: تابع بررسی شیب را پیاده سازی کنید. (برای محاسبه $\|x\|_2$ از تابع `np.linalg.norm(x)` کمک بگیرید).

مرحله سوم: بررسی شیب چند بعدی

مراحل قبل را به حالت چند بعدی تعمیم می‌دهیم:



تمرین سوم: تابع بررسی شیب را برای حالت چند بعدی مشابه قبل پیاده سازی کنید.