

تمرین کامپیوتری چهارم هوش مصنوعی

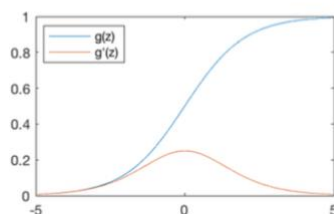
امیرمرتضی رضائی – 810003004

مرحله اول: پیاده سازی هر یک از لایه های یک شبکه عصبی

پیاده سازی موارد ذکر شده انجام شده و تست شد.

1- توابع فعالساز در شبکه های عصبی برای غیرخطی کردن عملیات شبکه و ایجاد قابلیت یادگیری توابع پیچیده تر به کار می روند. بدون استفاده از فعال سازها، یک شبکه عصبی حتی اگر لایه های زیادی داشته باشد، صرفاً یک مدل خطی خواهد بود که توانایی مدل سازی پیچیدگی های داده های واقعی را ندارد. برخی توابع فعالساز:

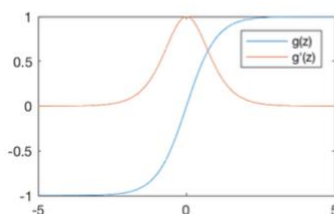
Sigmoid Function



$$g(z) = \frac{1}{1 + e^{-z}}$$

$$g'(z) = g(z)(1 - g(z))$$

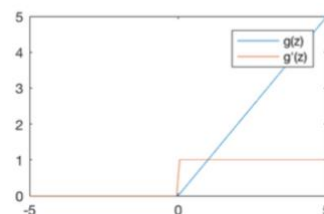
Hyperbolic Tangent



$$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

$$g'(z) = 1 - g(z)^2$$

Rectified Linear Unit (ReLU)



$$g(z) = \max(0, z)$$

$$g'(z) = \begin{cases} 1, & z > 0 \\ 0, & \text{otherwise} \end{cases}$$

2- تابع هزینه میزان خطای مدل را با مقایسه پیش‌بینی‌ها و مقادیر واقعی اندازه‌گیری می‌کند و هدف مدل در حین آموزش این است که این هزینه را به حداقل برساند.

- تابع هزینه MSE : برای مسایل رگرسیون به کار می‌رود و میزان تفاوت بین مقادیر پیش‌بینی شده مدل و مقادیر واقعی را اندازه‌گیری می‌کند:

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2$$

- تابع هزینه softmax : معمولا در مسائل دسته‌بندی چندکلاسه به کار می‌رود. این تابع برای تبدیل خروجی‌های مدل که معمولا پیش‌بینی‌های خام هستند، به احتمال‌های قابل تفسیر برای هر کلاس استفاده می‌شود.

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

3- این مبحث تدریس نشده است!

4- شبکه‌های عصبی در شرایطی مانند وجود داده‌های غیرخطی، داده‌های حجیم (BigData) و مسایل پیچیده دسته‌بندی و پیش‌بینی نسبت به الگوریتم‌های یادگیری ماشین کلاسیک برتری دارند.

5- در این حالت شبکه عصبی تلاش می‌کند که تمام ویژگی‌ها و تفاوت‌های موجود در داده‌های آموزشی را یاد بگیرد و به این ترتیب، برازش به داده‌ها به قدری دقیق می‌شود که مدل نمی‌تواند برای داده‌های جدید عمومی‌سازی کند. در واقع overfitting رخ داده و گویی نویز نیز مدل می‌شود!

مرحله دوم: کنار هم گذاشتن بخش‌ها در کنار هم

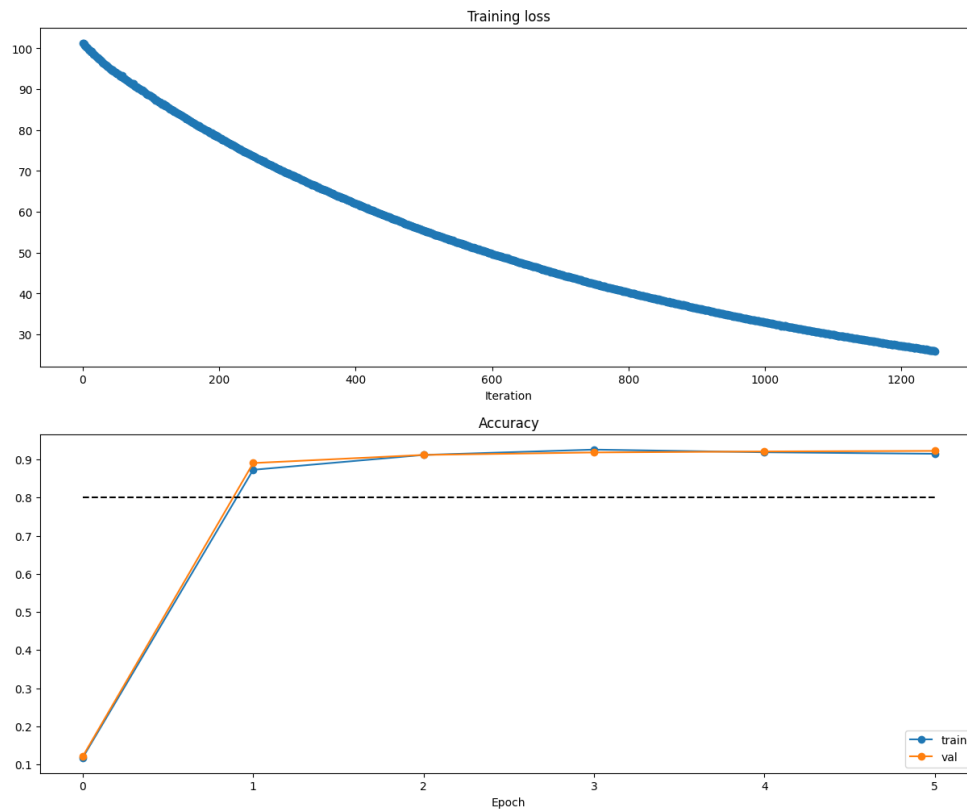
6- SGD: الگوریتم گرادیان کاهشی تصادفی یک روش بهینه‌سازی است که برای به‌روزرسانی وزن‌ها استفاده می‌شود. در این روش، به‌جای محاسبه گرادیان‌ها برای کل داده‌ها (که در گرادیان کاهشی معمولی رخ می‌دهد)، از یک نمونه تصادفی از داده‌ها برای به‌روزرسانی وزن‌ها استفاده می‌شود. بدین صورت که ابتدا وزن‌ها به‌صورت تصادفی یا با مقداری اولیه تنظیم می‌شوند؛ سپس برای هر نمونه تصادفی، گرادیان تابع هزینه نسبت به وزن‌ها محاسبه می‌شود. در نهایت نیز وزن‌ها به‌روزرسانی می‌شوند به‌طوری‌که مقدار تابع هزینه کاهش یابد.

$$w \leftarrow w + \alpha \cdot \nabla g(w)$$

Momentum: یک روش بهبود برای SGD است که به‌منظور کاهش نوسانات و تسریع همگرایی طراحی شده. این روش از شتاب در به‌روزرسانی وزن‌ها استفاده می‌کند که به‌صورت یک میانگین وزنی از گرادیان‌های قبلی ذخیره می‌شود. بدین صورت که به‌جای استفاده مستقیم از گرادیان‌ها، از یک شتاب یا سرعت استفاده می‌شود که ترکیبی از گرادیان‌های گذشته است. این سرعت به‌طور تدریجی با هر به‌روزرسانی وزن‌ها بروز می‌شود:

$$\begin{aligned} z &\leftarrow \beta \cdot z + \nabla g(w) \\ w &\leftarrow w + \alpha \cdot z \end{aligned}$$

مرحله سوم: آموزش و اعتبار سنجی شبکه عصبی ایجاد شده



همانطور که انتظار می‌رفت، با پیشرفت آموزش، مدل باید یاد بگیرد که پیش‌بینی‌های بهتری انجام دهد که باعث کاهش اتلاف می‌شود. این روند کاهشی در اتلاف نشان می‌دهد که مدل در حال بهبود است.

همچنین دقت باید به تدریج افزایش یابد زیرا مدل یاد می‌گیرد که داده‌ها را بهتر طبقه‌بندی کند.

در مقایسه با خط مرجع 80% نیز مشاهده می‌گردد که دقت اعتبارسنجی مدل بالای 80% باقی مانده و این نشان می‌دهد که مدل عملکرد خوبی دارد.

مرحله چهارم : آموزش و اعتبار سنجی شبکه عصبی پیچشی

7- این لایه‌ها به شبکه کمک می‌کنند تا ویژگی‌های محلی و الگوهای تکراری موجود در تصویر را شناسایی کنند. ویژگی‌هایی مانند لبه‌ها، بافت‌ها، اشکال هندسی و سایر خصوصیات بصری در تصاویر، به صورت خودکار توسط فیلترهای موجود در این لایه‌ها استخراج می‌شوند. در واقع لایه‌های پیچشی با استفاده از فیلترهایی که ابعاد کوچکی دارند به بخش‌های مختلف تصویر اعمال می‌شوند. این فیلترها قادرند ویژگی‌های محلی مانند لبه‌ها، رنگ‌ها یا الگوهای ساده را شناسایی کنند.

پس از اینکه ویژگی‌ها از تصویر توسط لایه‌های پیچشی استخراج شدند، نیاز است که این ویژگی‌ها به‌طور کامل ترکیب شوند تا یک نمای کلی از داده‌ها داشته باشیم. این کار توسط لایه‌های تماماً متصل انجام می‌شود در واقع این لایه‌ها می‌توانند ویژگی‌های استخراج شده از لایه‌های پیچشی را ترکیب کنند و به یک نمای برداری با ابعاد ثابت تبدیل کنند که مناسب برای پیش‌بینی یا هر نوع طبقه‌بندی است.

8- لایه‌های تماماً متصل هیچ‌گونه محلی‌سازی ندارند و نمی‌توانند ویژگی‌های محلی را به خوبی شبیه‌سازی کنند. همچنین آن‌ها برای هر نورون در لایه‌ی قبلی به یک وزن جداگانه نیاز دارند، که منجر به تعداد بسیار زیاد پارامترها می‌شود. ضمناً استفاده از لایه‌های تماماً متصل برای تصاویری با ابعاد بزرگ منجر به ایجاد یک ماتریس بسیار بزرگ از وزن‌ها می‌شود که آموزش آن بسیار دشوار خواهد بود و نیاز به حافظه بسیار زیادی دارد.