

Deep learning

HW 7

Babaksh Ansari
99527082

پیش از اینجا
99527082

اگر ایضاً تکرار هر کلمه در کلاس مربوطه می باشد می کنم.

کلمه	تکرار کل کلاس	تکرار کلاس	$p(w 0)$	$p(w 1)$	λ
نفاود	۲	۱	۰,۱	۰,۱۴	۰,۳۸
مذهبی	۲	۱	۰,۱	۰,۱۴	۰,۳۸
علی	۲	۱	۰,۱	۰,۱۴	۰,۳۸
ایرانی	۱	۲	۰,۱۱	۰,۲۱	-۰,۵۱
سیاسی	۱	۲	۰,۱۱	۰,۲۱	-۰,۵۱
گفتار	۱	۱	۰,۱۱	۰,۱۴	-۰,۱۷

$$\tau = 4$$

$$N(0) = 9$$

$$N(1) = 1$$

حال باید امثال این را با $p(w|\text{class})$ بفرمول کنیم.

$$p(w|\text{class}) = \frac{\text{freq}(w, \text{class})}{N_{\text{class}}}$$

و همچنان که بدانم سنتی دهنم بار هر کلمه مربوط به کلاس خواهد بود که از اینجا

$$\lambda = \log \frac{p(w|\text{class} = 0)}{p(w|\text{class} = 1)}$$

فرمول زیر استناده کنیم:

کلاس = ۱ داشته باشد و کلاس = ۰ نداشته باشد

نحوه هم: از آنجایی که مان است ماد داده است (الملحق) / بینم ال

در داده train وجود نداشته باشد / این کار بعثتی شود (متداول مابایی)

آن لیه ضغط شود و بینی خارخواب / کن از روش laplace smoothing

$$P(w|class) = \frac{\text{freq}(w, \text{class}) + i}{N_{\text{class}} + V_{\text{class}}} \quad \leftarrow \text{استنادی لینم} \text{ که کاری} \text{ ضریب} \alpha \text{ است}$$

تعادل نصیر فرد

حال طبق این فرمول ها جدول ضعفه قبل را برای لینم

حال همی تست ابتدا نیاز است که حفار logprior را ماسبه لینم این پرداخت

برای این است که مان است تعادل نهای در هر لایس بلکه بحسب و برای بقرا کا محاب

$$\log prior = \log \frac{D_{BOS}}{D_{EOS}} = 0 \quad \leftarrow \text{تعادل این عدد} / \text{این در احتمال ضرب لینم}$$

تعادل احتمال
د لایس صدر

حال همی محاسبه احتمال داده تست دیدیم:

$$\text{score} = \log prior + \lambda(w_1) + \dots + \lambda(w_n) \quad \leftarrow \text{کوادلاین} \text{ که} \text{ می} \text{ شود}$$

پنجمی داده سنت اول درایم :

$$\text{Score} = \log \text{Prior} + \lambda (\text{قیاده}) + \lambda (\text{خوبی}) + \lambda (\text{علو}) + \lambda (\text{اصحای})$$

$$\text{Score} = 0 + 0,13\Delta + 0,13\Delta + 0,13\Delta - 0,48 = 0,48 > 0$$

این داده سنت متعلق به لالس ضمیر است

پنجمی داده سنت دهم درایم :

$$\text{Score} = \log \text{Prior} + \lambda (\text{قیاده}) + \lambda (\text{خوبی}) + \lambda (\text{علو}) + \lambda (\text{اصحای}) + \lambda (\text{فرزشی})$$

در اینجا (فرزشی) ۰,۱۴ را نهایم میگیرد اکنون پنجم ضمیر است

$$\lambda (\text{فرزشی}) = \frac{0+1}{4} = 0,25$$

$$\text{Score} = 0 + 0,13\Delta + 0,13\Delta + 0,13\Delta - 0,48 + 0,25 = 0,48 > 0$$

این داده سنت متعلق به لالس ضمیر است

فرمول دارایم $- \log L(\theta)$ است چهارم

$$-\log L(\theta) = \sum_{i=1}^n (y_i \log \hat{y}_{\theta,i} + (1-y_i) \log (1-\hat{y}_{\theta,i}))$$

$$\hat{y}_{\theta,i} = \phi(x_i^T \omega)$$

له در این سوال داریم

سده با عتادم درودی

$$-\log L(\theta) = \sum_{i=1}^n [y_i \log (\phi(x_i^T \omega)) + (1-y_i) \log (1-\phi(x_i^T \omega))]$$

در نتیجه داریم :

۴- الف) از آنچه که این مدل دسای و امعی بسیار فعال خواهد بود و معمول است

همه ای آنکه در این مدل بتواند این ویژگی های غیرخطی را استخراج نمود از توابع activation

استناده ی لام

ب) از قطعه کد زیر که هر تابع لام خواهد و پیوسته را با عنوان تابع فعال سازی

استناده کرد (پس از پیوسته باشد) ، اما در عمل اگر تابع در نظر گرفته شود

باشد و در نظر گیری می شود یعنی نباشد چنان است که از دیگر خطا و مشکل لذت

و حتی اگر تابع یعنی هم نباشد ایجاد مشکلی ندارد

(الف) ۱)

بنابراین $f(x)$ هم معروف است و در قدرت رسمی (پیشگویی) Sigmoid

بن ۰.۵ بیاید. این تابع پیوسته، یکنواخت و متعادل است و منتهی آن

که بعد از این سه بدل است که تابع بله است. لیکن از مسئله ای که این است که در این

آن همراه باشد این است و مقدار این تابع از مقدار ایجاد شده در مسئله gradient descent

gradiente

را هم بوجود می آورد.

این تابع محدود است اما این صفر و ۱ بینیم که اندیشه ای تغیری:

تفاوت بسیار نسبت است و برای multi class است (استناده ی سه).

لیکن از حیثی که محدود است اینها را نمی توانند از این تابع خارج نمودند

لذا از این تابع خارج نمودند

ناهال NAHAL

از تابع هادر است ReLU و مقدار حقیقتی را صفر یا ۱ نمایش می‌کند. عواملی‌های آن را آنچه سریع آن است چون فرمول ساده‌ای دارد محاسبات آن راحت است. آنچه از خصلات آن هادر مستثنی نماید بود آن در نظره حضراست و همینطور هادر صفر باشد تمام مقاطعه می‌باشد و این باعث می‌شود همه نورون‌ها غیرفعال شوند. به همین دلیل از توابع ماتم ReLU با Leaky ReLU استفاده می‌گردد.

از تابع نزدیک sigmoid است با این تفاوت‌ها هم هادر و روکار Tanh بگذرد که اصولاً بسیاری از مانند آن است. این تابع در MNN ‌ها خصوصاً در لایه‌های مقادیر خوبی آن فقط نسبت نسبت و این باعث جلوگیری از خروج نیازگاری می‌شود و تغییر سریع آن را در لایه‌ها می‌گیرد.

چون خروجی ReLU ممکن باشد در آن Perrmann sigmoid نیز نیست. این باعث می‌شود فضای خروجی آن بزرگ تر و ببر شود اما تعداد واحد های عمل ممکن است ازین بود (با خاصیت در موافق حقیقتی)

(۶) همارا یک شبکه عصبی مصنوعی داشتیم که با ۴۰ نورون

است و یک لایه خروجی با ۳ نورون دلیل تعداد نورون‌های لایه اول پنهان‌تر از این

پنهان‌های ورودی است و تعداد نورون‌های لایه آخر پانزده برابر تعداد لایه‌های مiddenی است.

۳ است می‌باشد. یعنی نیم از نورون‌های لایه اول بستگی به نیم از نورون‌های داده‌ها داشتند.

داده‌وں چون اینجا ماتنها ۲۰ نورون بولی train داریم و داده‌های ما بعید نیست

کل الاین سیانی فراری کم که تعداد نورون های آن را تقریباً ۳۲ تا ۴۰ داشت.

ReLU سالیهای سیانی (الاین وروکی به الاین ۲) از activation function.

استفاده کی لئن چون سریع است ویرایی الاین خروجی چون لیع multi table classification.

از softmax استفاده کی لئن. پلی تابع loss نیز چون حستیندی که نم.

الاین لیع از Cross-Entropy استفاده کی لئن.

۶- خروجی Z حاوی در ReLU وارد می شود اما آن بزرگ تر رسانید
صفر خواهد شد.

حال طبق sigmoid اگر وروکی آن اعدام نیست باشد، خروجی آن بزرگ تر محسوب شود.
هر دو $f(x) = \frac{1}{1+e^{-x}}$ و $\text{ReLU}(x) = \max(0, x)$ هر دو خواهد بود.
همواره حقیر لایس دوم خواهد بود و بدل نیز تولید تعادل لایس اول را
رسانید لذت.

۷- الف) اصلی برنی نتایج deep learning و الکترونیک ML درین است

نه در ML نی توانست به صورت خود بخود تبیین ها و بازنگانی های جیورا

ایدیلیزرن و فقط درین مجموعه از تبیین های از سین تعریف شده جست و تجربه کرته

آن مدل‌های یادگیری عمیق بدون آنده مابه آن‌ها باید خودشان بازنایی‌های

جید را یاد می‌کنند و بهمین دلیل باید بیامشت پس از عملکرد پرکار است.

این درستگاهی ساده ۳۰ درصد.

ب) این سؤال هنگامی است و کامل است و اینسته به چندین حیاس مصلحت است.

برای کمال به معنی داده، تسلیخ مردم و بعدی خود مدل واسیه است

اما اگر نجوم در حالتی در تقطیع پیغام همچو داده ما بتوانیم آن را بازسازی کنیم در این

حای بالای پر است (با توجه به Vanishing gradient پس از تقویم پیوسم اگر مرده های

ما ساده و مستعد نتیجه لایه ۷ می‌توانند بهم باشد چون جمله است در این های عمیق

محل feature در این راسته ای را استخراج کنند شاید لایه ۱۱ پرکار باشد. اما اگر نیاز است که مدل

بعد از این کار استخراج کنند شاید لایه ۱۱ پرکار باشد. اما باز تا لایه ۷ می‌توانیم این را

نمایش و تجربه نلئم نمی‌توانیم جذب قطعی بگیریم.

۷) سپله های عمیق آر، نیزاً هرجچه مدل عینک آر باشد لایه های عمیقاً

می تواند با ترتیب دقیق تر کی تابع را پیشنهاد نماید

۸) خرابیا - لایه های بسیمی دوست Feature های بعدها برک را استخراج می کند

- می توانست داده های بعدها کم رو بسیر آنالیز کند
- می توانست روابط غیرخطی بسیر کرا استخراج کند

محاذب:

- اثر داده های ارزش کافی نباشد به دلیل فیلتر های غیر ضروری overfitting

نماینده

- حیزان هر نهاد محاذب را افزایش می دهد

- باید احتمالی vanishing gradient را شوند و نزول های لایه های ایتلی خوب باشند

خوب