

تحلیل حساسیت

سجاد سبزی
محمدرضا احمدی تشنیزی

۱۳ آذر ۱۴۰۲

تحلیل حساسیت (Sensitivity Analysis) در مدل‌های یادگیری ماشینی چیست؟

تحلیل حساسیت در زمینه مدل‌های یادگیری ماشینی به ارزیابی این موضوع می‌پردازد که چگونه خروجی یک مدل تحت تأثیر تغییرات متغیرهای ورودی قرار می‌گیرد. این یک تکنیک است که برای تعیین اینکه چقدر یک مدل نسبت به تغییرات در ویژگی‌های ورودی «حساس» است، استفاده می‌شود. این تحلیل می‌تواند در درک رفتار مدل تحت شرایط مختلف و شناسایی اینکه کدام ورودی‌ها بیشترین تأثیر را بر پیش‌بینی‌های مدل دارند، کمک کند.

هدف از تحلیل حساسیت

هدف اصلی تحلیل حساسیت، سنجش استحکام و قابلیت اطمینان مدل‌های یادگیری ماشینی است. این کمک می‌کند در شناسایی اینکه کدام متغیرهای ورودی تأثیرگذارترین هستند، درک رفتار مدل در سناریوهای مختلف، و در اشکال‌زدایی مدل با آشکار کردن وابستگی‌ها یا ناهنجاری‌های غیرمنتظره در پاسخ‌های مدل.

چه زمانی از تحلیل حساسیت استفاده می‌شود؟

تحلیل حساسیت در موارد زیر استفاده می‌شود:

- در فاز توسعه مدل، برای درک و بهبود عملکرد مدل.
- در مرحله تأیید مدل، برای ارزیابی استحکام مدل در برابر تغییرات در داده‌های ورودی.
- قبل از استقرار مدل، برای اطمینان از عملکرد خوب آن تحت شرایط واقعی پنهان مختلف.

مزایا و معایب

مزایا

- درک بهتر از رفتار مدل: تحلیل حساسیت به شما امکان می‌دهد تا درک کنید چگونه ورودی‌های مختلف بر خروجی‌های مدل شما تأثیر می‌گذارند. این برای مدل‌های پیچیده که رابطه بین ورودی و خروجی مستقیم نیست، حیاتی است.
- شناسایی ویژگی‌های کلیدی: این کمک می‌کند در شناسایی ویژگی‌های تأثیرگذار که تأثیر قابل توجهی بر پیش‌بینی‌های مدل دارند. این به ویژه در انتخاب ویژگی و درک اینکه کدام متغیرها برای بهبود مدل ارزش تمرکز دارند، مفید است.
- استحکام و قابلیت اطمینان مدل: با بررسی اینکه چگونه تغییرات در ورودی بر خروجی تأثیر می‌گذارد، می‌توانید استحکام مدل خود را ارزیابی کنید. این کمک می‌کند تا اطمینان حاصل شود که مدل به طور مداوم در برابر طیف وسیعی از مقادیر ورودی عمل می‌کند و به تغییرات کوچک بیش از حد حساس نیست.
- تسهیل در تنظیم و بهبود مدل: تحلیل حساسیت می‌تواند ضعف‌ها یا سوگیری‌های مدل را آشکار کند و توسعه‌دهندگان را در پالایش و بهبود معماری یا فرآیند آموزش مدل راهنمایی کند.
- کمک در فرآیند تأیید و تصدیق مدل: این ابزار ضروری در فرآیند تأیید مدل است و اطمینان می‌دهد که مدل تحت شرایط و سناریوهای ورودی مختلف همانطور که انتظار می‌رود رفتار کند.
- پشتیبانی از رعایت مقررات و مدیریت ریسک: در بخش‌هایی مانند مالی یا بهداشت، درک اینکه چگونه مدل به ورودی‌های مختلف واکنش نشان می‌دهد، برای مدیریت ریسک و برآوردن الزامات قانونی حیاتی است.

معایب

- پیچیدگی محاسباتی: برای مدل‌های بزرگ و پیچیده، تحلیل حساسیت می‌تواند از نظر محاسباتی گران و زمان‌بر باشد، که نیازمند منابع گسترده برای ارزیابی مدل در سناریوهای مختلف است.
- محدودیت‌ها در درک تعاملات پیچیده: برخی از روش‌های تحلیل حساسیت ممکن است به طور مؤثر تعاملات پیچیده و غیرخطی بین متغیرها را درک نکنند، به خصوص در مدل‌های بسیار پیچیده مانند شبکه‌های عصبی عمیق.
- امکان برداشت اشتباه: نتایج تحلیل حساسیت می‌توانند به اشتباه تعبیر شوند، به ویژه اگر فرضیات اصلی مدل یا روش تحلیل به درستی در نظر گرفته نشوند.
- وابستگی به کیفیت و دامنه داده‌ها: اثربخشی تحلیل حساسیت به شدت به کیفیت و دامنه داده‌های ورودی مورد استفاده بستگی دارد. اگر داده‌ها طیف کاملی از سناریوهای ممکن را نمایش ندهند، تحلیل ممکن است کاملاً قابل اعتماد نباشد.
- چالش‌ها در فضاهای با بعد بالا: در مدل‌هایی با تعداد بسیار زیادی از ویژگی‌ها (فضاهای با بعد بالا)، انجام تحلیل حساسیت جامع دشوارتر می‌شود و نتایج دشوارتر قابل تفسیر است.
- تشدید تعصب: اگر مدل یا داده‌ها دارای تعصب باشند، تحلیل حساسیت ممکن است به طور ناخواسته این تعصب‌ها را تشدید کند، به ویژه اگر با دقت انجام و تفسیر نشوند.

جنبه‌های اضافی

- کاربردها: تحلیل حساسیت به طور گسترده‌ای در زمینه‌هایی مانند مالی، بهداشت و درمان، و مدل‌سازی آب‌وهوایی استفاده می‌شود، جایی که درک تأثیر متغیرهای ورودی حیاتی است.
- روش‌ها: تکنیک‌هایی مانند روش سوبول (Sobol method)، تحلیل حساسیت مبتنی بر واریانس، یا رویکردهای ساده‌تر مانند تحلیل اختلال (perturbation analysis) به طور معمول استفاده می‌شوند.
- تلفیق با هوش مصنوعی قابل توضیح (XAI): تحلیل حساسیت اغلب جزئی از XAI است، که به روشن‌تر و قابل فهم‌تر شدن تصمیمات مدل کمک می‌کند.
- تأثیر پیش‌پردازش داده‌ها: نتایج تحلیل حساسیت می‌تواند تحت تأثیر نحوه پیش‌پردازش و نرمال‌سازی داده‌های ورودی قرار گیرد، که اهمیت دقت در آماده‌سازی داده‌ها را برجسته می‌کند.

کدهای PyTorch برای تحلیل حساسیت در GitHub

sensitivity_torch

- توضیحات: این بسته برای گرفتن مشتقات مرتبه اول و دوم از طریق بهینه‌سازی یا سایر فرایندهای نقطه ثابت طراحی شده است و بر روی PyTorch ساخته شده است.
- آدرس اینترنتی: https://github.com/rdyro/sensitivity_torch
- نصب: نصب با استفاده از pip با دستور `pip install git+https://github.com/rdyro/sensitivity_torch.git` یا از منبع.
- آزمایش: اجرای تمام تست‌های واحد با استفاده از `python3 test. setup.py`

SA_DNN: تحلیل حساسیت شبکه‌های عصبی عمیق

- توضیحات: ارائه اسکریپت‌های پایتون برای مجموعه داده‌های CIFAR10 و MNIST، شامل پیاده‌سازی شبکه‌هایی مانند ResNet50 و DenseNet121.
- آدرس اینترنتی: https://github.com/shu-hai/SA_DNN
- استفاده: شامل اسکریپت‌های مختلف پایتون برای مدیریت مجموعه داده‌ها و معماری‌های شبکه مختلف، مانند `CIFAR10_DenseNet121.py` و `MNIST_ResNet50.py`.

torchsde: حل‌کننده‌های SDE متفاوت با پشتیبانی GPU و تحلیل حساسیت مؤثر

- توضیحات: پیاده‌سازی PyTorch از حل‌کننده‌های معادلات دیفرانسیل تصادفی (SDE) متفاوت، با پشتیبانی GPU و تحلیل حساسیت مؤثر.
- آدرس اینترنتی: <https://github.com/google-research/torchsde>
- نصب: نصب از طریق pip با استفاده از `pip install torchsde`. نیازمند Python ≥ 3.8 و PyTorch $\geq 1.6.0$.