

موقعیت Outliers در توزیع لاپلاس

بررسی مقایسه‌ای با توزیع نرمال

تحلیل آماری

۲۲ شهریور ۱۴۰۴

۱ مقدمه

سوال اصلی این است که outlierها در توزیع لاپلاسی در کجا قرار دارند و چرا؟

پاسخ کلی

در توزیع لاپلاسی، outlierها کمتر مشهود هستند و در واقع این توزیع های outlier کمتری نسبت به توزیع نرمال تولید می‌کند.

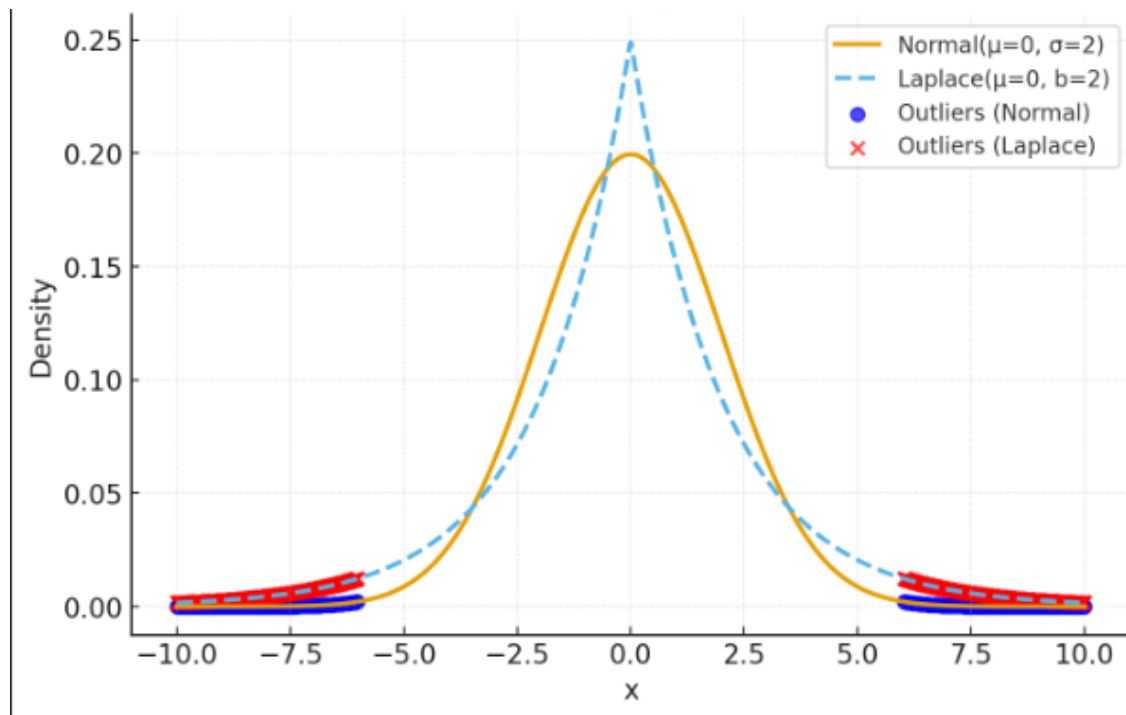
۲ موقعیت Outliers در توزیع لاپلاسی

۱.۲ مقایسه با توزیع نرمال

- نقاط آبی (Outliers Normal): در مناطق نزدیک‌تری به مرکز قرار دارند (حدود $x \approx \pm 4$)
- نقاط قرمز (Outliers Laplace): در مناطق دورتری از مرکز قرار دارند (حدود $x \approx \pm 7$ یا ± 8)

مشاهدات کلیدی از نمودار

- توزیع نرمال (خط زرد): دم‌های نازک‌تر دارد، بنابراین مقادیری که در $x \approx \pm 4$ قرار دارند به سرعت به outlier تبدیل می‌شوند
- توزیع لاپلاس (خط آبی چین‌دار): دم‌های پهن‌تر دارد، بنابراین برای اینکه یک نقطه outlier محسوب شود باید خیلی دورتر از مرکز باشد



شکل ۱: مقایسه موقعیت outliers در توزیع نرمال و لاپلاس. نقاط آبی (outliers نرمال) نزدیک‌تر به مرکز و نقاط قرمز (outliers لاپلاس) دورتر از مرکز قرار دارند. این نمودار نشان می‌دهد که توزیع لاپلاس (خط آبی چین‌دار) دارای دم‌های سنگین‌تر نسبت به توزیع نرمال (خط زرد) است.

۳ دلایل این تفاوت

۱.۳ Tails Heavy (دم‌های سنگین)

توزیع لاپلاسی دارای دم‌های سنگین‌تری نسبت به توزیع نرمال است. این یعنی:

- مقادیر دورتر از مرکز احتمال بیشتری دارند که رخ دهند
- آنچه در توزیع نرمال outlier محسوب می‌شود، در لاپلاسی ممکن است طبیعی باشد

۲.۳ شکل توزیع

توزیع لاپلاسی دارای ویژگی‌های زیر است:

- قله تیزتر (Sharper) (Peak): نسبت به توزیع نرمال
- دم‌های آهسته‌تر: به آرامی کاهش می‌یابند
- ساختار مقاوم: مقادیر افراطی بیشتر در محدوده قابل قبول قرار می‌گیرند

۳.۳ تحلیل ریاضی

تابع چگالی احتمال توزیع‌ها:
توزیع نرمال:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

توزیع لاپلاس:

$$f(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$

تفاوت کلیدی در نحوه کاهش تابع در دم‌هاست:

- نرمال: کاهش نمایی با توان دوم (e^{-x^2})
- لاپلاس: کاهش نمایی خطی ($e^{-|x|}$)

۴ کاربردهای عملی

۱.۴ کاربردهای توزیع لاپلاسی

توزیع لاپلاسی به دلیل مقاومت در برابر outlier در موارد زیر استفاده می‌شود:

۱. Statistics Robust: وقتی می‌خواهیم نسبت به outlier مقاوم باشیم

۲. Regression Lasso: در یادگیری ماشین برای تنظیم‌سازی

۳. پردازش سیگنال: جایی که نویز دارای outlier است

۴. تحلیل تصاویر: برای حذف نویزهای نقطه‌ای

۵. آمار بیزی: به عنوان prior مناسب

۶. مدل‌سازی مالی: برای بازارهای ناپایدار

۵ نتیجه‌گیری

خلاصه پاسخ

موقعیت outlierها در توزیع لاپلاسی:

- در مناطق دورتری از مرکز نسبت به توزیع نرمال قرار می‌گیرند
- به دلیل heavy tails، کمتر به عنوان outlier شناخته می‌شوند
- این ویژگی باعث می‌شود لاپلاس برای داده‌های دارای نویز یا outlier طبیعی مناسب‌تر باشد

این تحلیل نشان می‌دهد که انتخاب توزیع مناسب بر اساس ماهیت داده‌ها و نحوه برخورد مطلوب با outlierها بسیار مهم است.