

روش Bootstrap در آمار و احتمالات استفاده می‌شود و برای تخمین توزیع نمونه‌ای یک آماره استفاده می‌شود.

- روش Bootstrap اولین بار توسط برادلی افریم در سال ۱۹۷۹ معرفی شد و او این نام را به آن داد. برای توضیح دلیل انتخاب این نام، باید به معنی واژه bootstrap در زبان انگلیسی مراجعه کنیم. در زبان انگلیسی، bootstrap به معنای چیزی است که با استفاده از خودش راه اندازی می‌شود یا به طور ساده‌تر، یک فرآیند شروع می‌شود. این اصطلاح معمولاً در مواردی استفاده می‌شود که یک عملیات یا فرآیند نیاز به خودبیلانسی یا استقلال دارد. برادلی افریم، در معرفی روش Bootstrap، تصویری را ایجاد کرد که این روش از خودش شروع می‌شود و بدون نیاز به فرضیه‌های خاص، توزیع نمونه‌ای را تخمین می‌زند. با توجه به این تصویر و خاصیت خودبیلانسی و استقلال روش، او به آن نام Bootstrap را داد. بنابراین، نامگذاری Bootstrap برای این روش نسبتاً تصادفی بوده و معنای واژه در زبان انگلیسی با خاصیت‌های روش Bootstrap سازگاری دارد.
- مزیت اصلی روش Bootstrap در این است که اجازه می‌دهد تا بدون فرضیه‌های آماری خاص، تخمین‌های نمونه‌ای از ویژگی‌های یک جمعیت را محاسبه کند. برخلاف روش‌های سنتی که نیاز به فرضیه‌های قوی و پیچیده‌ای دارند، Bootstrap بر اساس توزیع نمونه‌های جدید، توزیع نمونه‌ای مورد نظر را تقریب می‌زند.
- مزایای دیگر روش Bootstrap شامل موارد زیر می‌شود:

انعطاف‌پذیری: روش Bootstrap قابل استفاده در تخمین هر نوع آماره است و به آن پارامتریک یا غیرپارامتریک نیست.

استفاده آسان: اجرای روش Bootstrap نسبتاً ساده است و نیازی به فرمول‌های پیچیده ریاضی ندارد.

استفاده از داده‌های واقعی: در Bootstrap، از داده‌های واقعی به عنوان نمونه استفاده می‌شود و نیازی به فرضیه‌های خاص درباره توزیع اصلی داده‌ها نیست.

مقاومت در برابر توزیع غیرنرمال: روش Bootstrap برای تخمین توزیع نمونه‌ای استفاده می‌شود و بنابراین نسبت به فرضیه‌هایی مانند توزیع نرمالیتی داده‌ها مقاوم است.

- با این حال، روش Bootstrap همچنین معایبی نیز دارد:

مصرف منابع محاسباتی: استفاده از Bootstrap ممکن است نیاز به محاسبات زمان‌بر داشته باشد، زیرا برای ایجاد نمونه‌های جدید باید عملیات تکرار شود.

وابستگی به داده‌های موجود: روش Bootstrap بر اساس داده‌های موجود استوار است. بنابراین، اگر داده‌های اصلی تغییر کنند، تخمین‌های Bootstrap نیز تغییر می‌کنند.

برخی حالات استثنا: در برخی موارد خاص، مانند داده‌های با انحراف معمول بسیار بالا یا توزیع‌هایی با دم بلند، روش Bootstrap ممکن است عملکرد نامناسبی داشته باشد.

در کل، روش Bootstrap یک ابزار قدرتمند در تخمین توزیع نمونه‌ای است که امکان تحلیل داده‌های واقعی را بدون نیاز به فرضیه‌های خاص در اختیار محققان قرار می‌دهد. با این حال، محدودیت‌ها و معایب آن را نیز در نظر بگیرید و در تفسیر نتایج به دقت و احتیاط عمل کنید.

شیوه ارزیابی Bootstrap0.632 یک روش ارزیابی در مدلسازی آماری است که برای تخمین عملکرد یک مدل استفاده می‌شود. این روش معمولاً در مسائل پیش‌بینی و ارزیابی مدل‌های آماری مورد استفاده قرار می‌گیرد.

شیوه Bootstrap0.632 براساس روش Bootstrap است و از ترکیب دو معیار، یعنی عملکرد دقت (accuracy) و خطا در نمونه‌های تست (out-of-sample error)، برای ارزیابی مدل استفاده می‌کند.

مراحل اجرای روش Bootstrap0.632 به شرح زیر است:

۱. از داده‌ها، یک نمونه بازیابی شده تصادفی برداری می‌شود (با جایگذاری)، که تعداد نمونه‌های آن با تعداد داده‌های اصلی برابر است. این نمونه مورد استفاده برای ساخت مدل مورد نظر قرار می‌گیرد.

۲. مدل مورد نظر روی نمونه بازیابی شده آموزش داده می‌شود و با استفاده از داده‌های بازیابی شده، پیش‌بینی می‌شود.

۳. مدل آموزش دیده روی داده‌های بازیابی شده ارزیابی می‌شود و عملکرد دقت آن محاسبه می‌شود.

۴. سپس، مدل آموزش دیده روی داده‌های اصلی (غیر بازیابی شده) ارزیابی می‌شود و خطای نمونه خارج از مجموعه آموزش (out-of-sample error) محاسبه می‌شود.

۵. در نهایت، معیار Bootstrap0.632 برابر است با  $0.632 \times$  ضرب شده در عملکرد دقت مدل روی داده‌های بازیابی شده و  $0.368 \times$  ضرب شده در خطای نمونه خارج از مجموعه آموزش مدل روی داده‌های اصلی. این معیار به عنوان یک تخمین برازش مناسب برای عملکرد مدل استفاده می‌شود.

شیوه Bootstrap0.632 به دلیل ترکیبی از دقت و خطای نمونه خارج از مجموعه آموزش، تلاش می‌کند تا برازش مدل را در نمونه‌های تست بهبود دهد و از مشکلاتی مانند بیش‌برازش (overfitting) در مدل‌های آماری جلوگیری کند.