

انتخاب متغیر با استفاده از زنجیر مارکف برگشت پذیر

علی آقامحمدی، عضو هیأت علمی گروه آمار، دانشگاه زنجان

فاطمه نجفی قراجه، دانشجوی کارشناسی ارشد، دانشگاه زنجان

چکیده: انتخاب متغیر، یکی از مهم ترین موضوعات در مدل سازی آماری محسوب می شود که کاربرد وسیعی در تحلیل های آماری دارد. بسیاری از روش های انتخاب متغیر با استفاده از مدل رگرسیون خطی انجام می شوند. در بین مدل های رگرسیون خطی، مدل رگرسیون تاوانیده از اهمیت خاصی برخوردار است و مدل رگرسیونی لاسو کاربردی ترین مدل رگرسیون تاوانیده است که در سال های اخیر از دیدگاه آمار بیزی ارائه شده و مورد بررسی قرار گرفته است. در این مقاله، از دیدگاه آمار بیزی، با تعریف توزیع پیشینی روی تعداد پارامترهایی که در مدل تأثیر گذارند، مدل رگرسیونی با تاوان لاسو پیشنهاد می شود. چون توزیع پسینی پارامترها شکل بسته ای ندارد، برای استنباط از روش های عددی زنجیر مارکف برگشت پذیر استفاده می شود. شبیه سازی ها نشان می دهند که مدل ارائه شده در مقایسه با سایر روش ها خطای برآورد کمتری دارد.

کلمات کلیدی: انتخاب متغیر، مدل رگرسیونی با تاوان لاسو، مدل بیزی، روش زنجیر مارکف برگشت پذیر

مقدمه

که در آن $X_{n \times p}$ ماتریس طرح، $y_{n \times 1}$ بردار مشاهدات به طول n ، $\beta_{p \times 1}$ بردار ضرایب به طول p و $e_{n \times 1}$ بردار خطاها می باشد که مؤلفه های آن مستقل و دارای توزیع نرمال با میانگین صفر و واریانس ثابت σ^2 هستند، یعنی $e_{n \times 1} \sim N_n(0, \sigma^2 I)$.

برآوردگر معمول پارامتر β ، برآوردگر کمترین مربعات خطا می باشد. اما در استفاده از این مدل، مشکل قابل توجه، دشواری برازش مدل به داده هایی است که تعداد مشاهدات در آنها کم و تعداد پیشگوها زیاد است. یکی از روش های مقابله با این مشکل، استفاده از روش های انتخاب متغیر است. در آمار فراوانی گرا و آمار بیزی روش های متعددی برای انتخاب زیرمجموعه ای از متغیرهای پیشگو که بیشترین تأثیر

پیشرفت روزافزون علوم مختلف و به وجود آمدن تکنیک های جدید این امکان را فراهم کرده است تا بتوان حجم وسیعی از داده ها در مورد متغیرهای مختلف به دست آورد. وجود این داده های حجیم در اکثر شاخه ها، از مطالعات بهداشتی گرفته تا مدیریت ریسک، تفکر آماری را به طور کلی دگرگون کرده است. در این گونه موارد انتخاب متغیر، خود را به عنوان یک ابزار بسیار توانمند و حیاتی برای زمینه های اکتشافی در علوم مختلف مطرح کرده است. مدل رگرسیون خطی چندگانه زیر را در نظر بگیرید:

$$y = X\beta + e \quad (1)$$