

تمرین 11: مدل رگرسیون

مهلت:

1402/06/08, 12:00 (30 AUGUST 2023)

نمره کل تمرین: 1200

دکتر مسعود کاویانی

مدرسه دیتاساینس فیلگر

مجموعه داده‌ی $HR.CSV$ در ضمیمه را در نظر بگیرید. هر سطر از این مجموعه‌ی داده یک کارگر در کارخانه‌ی تولیدی را نمایش می‌دهد که دارای ویژگی‌هایی است. ستون $LEFT$ نشان می‌دهد که آن کارگر استعفا داده است یا خیر (۱ یا ۰). می‌خواهیم با استفاده از یک الگوریتم رگرسیون و استفاده از مجموعه‌ی داده‌ی $HR.CSV$ یک مدل ایجاد کنیم که بتواند احتمال استعفای یک شخص را با استفاده از ستون $LEFT$ و بقیه‌ی ستون‌ها (که متغیرهای مستقل هستند) پیش‌بینی کند.

توجه کنید که ستون $LEFT$ دو مقدار صفر و یک را دارد که اگر به عنوان متغیر هدف به یک الگوریتم رگرسیون داده شود، الگوریتم حالت بین صفر و یک (یعنی احتمالی) را می‌تواند یاد بگیرد.

سعی کنید از الگوریتم‌های رگرسیون داخل $SKLEARN$ به همراه $GRIDSEARCH$ برای پیدا کردن بهترین ابرپارامترها کمک بگیرید.

