

بسمه تعالی

دوره یادگیری ماشین نظارت شده
تمرین اول: پیش پردازش و رگرسیون

شرح تمرین

یک مجموعه داده به نام *CSM_dataset.xlsx* مربوط به یک سایت نقد فیلم در پوشه دیتاست موجود می باشد.

الف) در ابتدا مجموعه داده را بخوانید.

ب) پیش پردازش های لازم را بر روی مجموعه داده انجام داده و آن را به دو بخش آزمون و آموزش به نسبت ۲۰:۸۰ تقسیم نمایید.

ج) رگرسیونی مبتنی بر نزول در راستای گرادیان برای تخمین ستون *Gross* آموزش دهید.

د) خطای مربوط به *MSE* را گزارش دهید.

ه) آیا می توان ستون مربوط به نام فیلم را حذف کرد؟ چرا؟