

دانشگاه اصفهان - رشته علوم کامپیوتر
یادگیری ماشین

تمرین ۱: مدل‌های خطی و روش‌های ارزیابی آنها

شماره دانشجویی:

نام و نام خانوادگی:

پرسش ۱.

می‌دانیم پس از آن‌که یک مدل یادگیرنده مرحله آموزش خود را سپری کرد، باید توسط چندین معیار، ارزیابی و سنجیده شود. همچنین اگر مسئله موردنظر از نوع رگرسیون باشد، خروجی مدل یک بردار مانند \hat{Y} و مقادیر برچسب نیز درون بردار دیگری مانند Y قرار دارند که همگی از نوع اعداد پیوسته هستند.

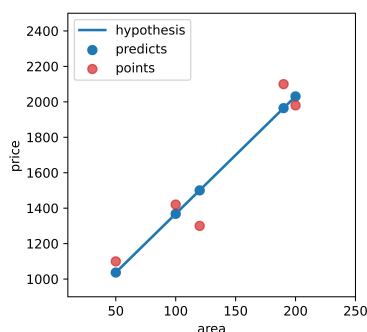
$$Y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \hat{Y} = \begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix}$$

به عنوان مثال یکی از معیارهای ارزیابی مدل‌های رگرسیونی، میانگین مربع خطا^۱ می‌باشد که به صورت زیر بیان می‌شود:

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (۱)$$

الف. مجموعه داده‌ای از مساحت خانه و قیمت آن به همراه پیش‌بینی مدل به شما داده شده است (جدول ۱). در شکل ۱ نقاط این مجموعه داده با در نظر گرفتن مساحت خانه در محور x و قیمت آن در محور y به همراه فرضیه h که یک خط تخمین زده شده توسط الگوریتم رگرسیون خطی است، نشان داده شده است. باتوجه به خط فرضیه تخمین زده شده، هر نقطه از این مجموعه داده که به خط فرضیه داده شود، دارای خطای e_i است. مقدار MSE را برای این مجموعه داده محاسبه کرده و مقدار حاصل از آن را توصیف کنید و توضیح دهید که این معیار ارزیابی چه معایب و مزایایی دارد. (می‌توانید با ارائه یک مثال، توضیحات کامل‌تری ارائه دهید)

^۱Mean Square Error



شکل ۱: نمایش دو بُعدی داده‌ها به همراه فرضیه تخمین زده شده. برای اطلاعات بیشتر و کد زدن با این مجموعه داده، می‌توانید از کدی که در [این] آدرس نوشته شده است استفاده کنید.

جدول ۱: مجموعه داده قیمت خانه‌ها

قیمت پیش‌بینی شده	قیمت واقعی	مساحت خانه
۱۹۶۴	۲۱۰۰	۱۹۰
۲۰۳۱	۱۹۸۰	۲۰۰
۱۳۶۸	۱۴۲۰	۱۰۰
۱۰۳۷	۱۱۰۰	۵۰
۱۵۰۰	۱۳۰۰	۱۲۰

ب. دو معیار ارزیابی دیگر برای مسائل رگرسیون انتخاب کرده و مقدار آن‌ها را برای مجموعه داده جدول ۱ محاسبه کنید. همچنین توضیح دهید که هرکدام از چه رابطه‌ای بدست می‌آیند و معایب و مزایای آن‌ها نسبت به یکدیگر چگونه است.

پاسخ.

الف.

ب.

پرسش ۲.

فرض کنید یک مجموعه داده دارای دو یا چند برچسب باشد و قرار است یک مدل یادگیرنده، مقدار آن‌ها را پیش‌بینی کند. چه راهکاری برای حل چنین مسئله‌ای ارائه می‌دهید؟ (به عنوان مثال، فرض کنید که قرار است در یک مسئله تشخیص اشیا در پردازش تصویر، تمامی حیوانات در تصویر را تشخیص داده و دور تصویر آن حیوانات، یک باکس مستطیل مانند رسم کنید. برای این کار نیاز است که مختصات دو نقطه از این مستطیل را در صفحه پیش‌بینی کنید که هرکدام دارای مؤلفه x و y هستند)

پاسخ.