

تمرین سری اول درس یادگیری ماشین – رگرسیون غیر خطی

در این تمرین، از شما خواسته شده است تا کد مربوط به رگرسیون غیر خطی تک متغیره را تکمیل کنید. برای این کار، نیاز است تا شما در محاسبات، توان‌های چند جمله‌ای متغیر ورودی را، تا مرتبه k حساب کرده و به عنوان یک ستون جدید، به دادگان خود اضافه کنید. همچنین به عنوان دادگان مورد نظر، از داده‌های خطی (مشابه دادگان موجود در کد اول رگرسیون خطی) استفاده کنید. برای انجام این تمرین، باید نمودارهای زیر را رسم کرده و سپس به سوال‌های مطرح شده، جواب دهید:

۱. نمودار خط نهایی مدل برای مجموعه دادگان با تعداد ۵، ۱۰، ۲۵ و ۱۰۰ نقطه داده و تعداد توان‌های چند جمله‌ای $k = 1$

۲. نمودار خط نهایی مدل برای مجموعه دادگان با تعداد ۵، ۱۰، ۲۵ و ۱۰۰ نقطه داده و تعداد توان‌های چند جمله‌ای $k = 4$

۳. نمودار خط نهایی مدل برای مجموعه دادگان با تعداد ۵، ۱۰، ۲۵ و ۱۰۰ نقطه داده و تعداد توان‌های چند جمله‌ای $k = 16$

۴. نمودار خطای نهایی مدل پس از آموزش برای مجموعه دادگان با تعداد ۵، ۱۰، ۲۵ و ۱۰۰ نقطه داده و تعداد توان‌های چند جمله‌ای از ۱ تا ۱۰ (محور عمودی نمودار را برابر با خطای MSE قرار دهید و محور افقی نیز تعداد توان‌های چند جمله‌ای باشد. سپس چهار خط با رنگ‌های متفاوت رسم کنید که هر خط نشان دهنده اطلاعات مربوط به یکی از مجموعه دادگان باشد)

* نکته: در رسم نمودارها، می‌توان به تابع `plot` یک آرگومان رشته داد که ویژگی‌های نمودار را مشخص می‌کند. این رشته می‌تواند دارای حروف `r,g,b,k,y,c,m` باشد که رنگ نمودار را مشخص می‌کنند (به ترتیب از راست به چپ: ارغوانی، فیروزه‌ای، زرد، مشکی، آبی، سبز، قرمز) همچنین این رشته می‌تواند شامل نمادهای `o`، `-o`، `-`، ... باشد که نوع خط را مشخص می‌کنند.

* نکته: در رسم نمودارها، می‌توان به تابع `plot` یک آرگومان `label` پاس داد که در آن، برای نمودار رسم شده، یک اسم مشخص کرد. سپس با اجرای دستور `plt.legend()` می‌توان نام‌ها را نمایش داد.

سوالاتی که در مورد آنها فکر کنید و تحلیل خود را بنویسید:

۱. به نظر شما برای دادگان خطی، استفاده از رگرسیون‌های غیر خطی نیز می‌تواند دقت کافی را کسب کند؟

۲. اگر هیچ دانشی در مورد دادگان خود نداشته باشیم، چطور می‌توانیم توان چند جمله‌ای لازم را بیابیم؟

۳. آیا کاهش خطای مدل بر روی مجموعه دادگان اولیه، همواره مطلوب است؟ آیا ممکن است مدلی که در دادگان آموزشی خطای کمتری داشته، در عمل بدتر عمل کند؟

۴. چطور می‌توانیم مشکل ایجاد شده برای مدل‌های چند جمله‌ای با توان بالا را حل کنیم؟

تاریخ تحویل تمارین: جمعه ۱۷ مرداد. برای تحویل نیز، کدهای خود را به همراه نمودارها در گیت‌هاب خود بارگزاری کرده و لینک آن را به همراه پاسخ سوالات به ایمیل بنده ارسال کنید. عنوان ایمیل را به صورت Name-ML-T۱ قرار دهید.
mohamadqodosi۱۹۹۶@gmail.com