

تمرین سری دوم درس یادگیری ماشین

Overfitting – Underfitting

در این تمرین می‌خواهیم رفتارهای **overfitting** و **underfitting** را در یک مدل ایجاد کنیم و نتایج آن را مشاهده کنیم تا با این پدیده بیشتر آشنا شویم. برای انجام این کار، به عنوان مجموعه دادگان، یک مجموعه داده با استفاده از یک معادله درجه چهارم دلخواه ایجاد کنید (برای این منظور، کافیست زمان محاسبه y توان دوم، سوم و چهارم x را نیز دخیل کنید). این مجموعه داده، باید دارای ۲۰۰ نمونه باشد. سپس با تکنیک بیان شده، دادگان خود را به دو بخش **train** و **test** تقسیم کنید (لزومی به ایجاد داده **validation** نیست) سپس ۵ خط با توان چند جمله‌ای ۱، ۲، ۴، ۸ و ۱۶ را به این مجموعه دادگان فیت کنید و سپس اقدام به بررسی **overfitting** و **underfitting** در این مدل کنید.

نمودارهایی که باید تولید شوند:

1. نمودار منحنی به دست آمده برای هر یک از خطوط به همراه مجموعه دادگان در رسم این نمودار، دادگان **test** با نقاط فیروزه‌ای (C') و دادگان **train** با نقاط آبی مشخص شوند.
2. نمودار خطای منحنی‌های نهایی برای دادگان **train** و **test** به تفکیک به ازای توان چند جمله‌ای (محور افقی توان چند جمله‌ای و محور عمودی خطای **MSE** را نشان می‌دهد).
3. نمودار **learning curve** به سبک قدیمی برای هر یک از توان‌های چند جمله‌ای (مشابه نمودار صفحه ۱۳ PDF چهارم)

* برای رسم نمودار یادگیری، می‌توانید از تابع **learning_curve** در کتابخانه **sklearn** کمک بگیرید.

همچنین انتظار می‌رود پس از انجام این تمرین، در مورد نتایجی که می‌توان از نمودارها گرفت بحث کنید و **overfitting** و **underfitting** را در تمام مدل‌ها به کمک نمودارها بررسی و اثبات کنید. همچنین با کمک **learning curve** بررسی کنید که به نظر شما، در هر یک از مدل‌ها، چقدر داده برای آموزش مدل کافیهست و آیا جمع آوری دادگان بیشتر کمکی به ما می‌کند یا خیر؟

برای تحویل این تمرین، کدهای خود را در گیت‌هاب خود قرار دهید و فایل گزارش را به همراه آدرس ریپازیتوری برای بنده ارسال کنید. همچنین عنوان ایمیل خود را، **Name-ML-T2** قرار دهید.

mohamadqodosi1996@gmail.com

تاریخ تحویل : جمعه ۲۴ مرداد