

پرسش های کلاسی

سری شماره دو

تمرین ۱. با توجه به داده های پوشه Q1S2 به سوالات زیر پاسخ دهید. برای نوشتن تابع `evaluation` از نمونه قالب این پوشه استفاده کنید.

`Data1` تولید شده از ۱۰ نقطه تصادفی برای تابع $y = \sin(2\pi x)$ در بازه $[0, 1]$ است.

`Data2` با اضافه کردن یک نویز به t_i های `Data1` تولید شده اند.

`Data3` شامل ۵۰ نقطه تصادفی همانند `Data1` است. از این داده به عنوان داده تست در سوالات استفاده شود.

`Data4` این داده با در نظر گرفتن دو داده پرت در `Data2` تولید شده است.

`Data5` مشابه ساختار `Data2` تولید شده و تنها تعداد بیشتری نمونه دارد.

آ) برای `Data1` تا `Data5` به جز `Data3` با درجات $M = 1:9$ ، چند جمله ای حاصل از رگرسیون را به دست آورده و با استفاده از تابع `evaluate` دقت را برای داده های آموزشی متناظر خود (داده های آموزشی) حساب کنید. (برای M های مختلف).

ب) نتیجه این مدل ها را برای داده های تست `Data3` حساب کرده و دقت را برای M های مختلف به تفکیک روش گزارش کنید. همچنین برای هر M ، ماتریس ضرایب $(\Phi^T \Phi)$ را تشکیل داده و با استفاده از دستور `numpy.linalg.cond` عدد حالت را حساب کنید. عدد حالت ها را در یک جدول آورده و همینطور رسم کنید.

نکته: دقت کنید در اینجا مدل رگرسیون خطی به صورت

$$\min \|\Phi \omega - t\|_2$$

خواهد بود که جواب آن از حل $(\Phi^T \Phi) \omega = \Phi^T t$ به دست می آید. نیازی به حل این دستگاه نیست و از کتابخانه `sklearn` برای حل رگرسیون چنین دستگاهی استفاده کنید.

ج) در مورد ارتباط بین `RMSE` های گزارش شده برای داده های آموزشی و تست و M های مختلف و عدد حالت ماتریس ضرایب، تحلیل ارائه دهید.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \tilde{y}_i|, \quad RMSE = \sqrt{\frac{\sum_{i=1}^n (\tilde{y}_i - y_i)^2}{n}}$$

دوره یادگیری ماشین / دانشگاه تربیت مدرس / گروه علوم کامپیوتر / نیمسال اول سال تحصیلی ۱۴۰۱-۱۴۰۲
استاد درس : دکتر منصور رزقی آهق (Mrezghi.ir)