

بسم الله الرحمن الرحيم

یادگیری ماشین

تمرین دوم

تهیه شده توسط:

فاطمه ساوینی

میلاد مسرور

در مدلی که برای **fit** کردن روی معادله درجه چهار، از خط با توان چند جمله ای یک استفاده کردیم، دچار **underfitting** شده ایم و این به خاطر ساده بودن بیش از حد مدل آموزش داده شده است. از **learningcurve** این مدل خاص میتوان **underfitting** بودن را توجیه کرد. مقدار **accuracy(acc)** بر دادگان **train** پایین تر از حد معمول است.

برای مدل با خط با توان چند جمله ای ۲، با توجه به **learningcurve** میتوان به این نتیجه رسید که مدل نه دچار **overfitting** شده و نه دچار **underfitting**. اما به خاطر **MSE** بالایی که برای دادگان **Train** و **Test** گزارش شده است، مدل مناسبی نمی باشد.

برای مدل با خط با توان چند جمله ای ۴ و ۸ با توجه به **learningcurve**، مدل دچار هیچ گونه مشکلی نمی باشد و همچنین این دو مدل خاص دارای **MSE** پایینی هستند، بنابراین برای دادگان ما این دو مدل خیلی خوب کار میکنند.

برای مدل با خط به توان چند جمله ای ۱۶، **learningcurve** ما مقادیر **acc** منفی داده است که باعث تعجب ما شده است و حقیقتاً نظر خاصی در این مورد نداریم.

با توجه به نمودار **MSE vs The power of polynomials** برای این دادگان مشخص توان های ۴ و ۸ برای چند جمله ای های مورد استفاده برای مدل کافی میباشد. افزایش توان چند جمله ای جز بار محاسباتی زیاد، فایده ای برای ما ندارند.

برای مدل با خط به توان چند جمله ای ۱ ، با توجه به **learning curve** در **training set size** برابر با ۳۰ به نتایج خوبی میرسیم اما در هر صورت مدل ما **underfitting** میباشد.

برای توان ۲ هم **training set size** برابر با ۳۰ مناسب میباشد اما در هر صورت در این مدل ما **MSE** بالایی داریم.

برای توان های ۴ و ۸ ، با توجه به **learning curve** آن ها ، **training set size** برابر با ۶۰ و تقریبا ۱۸ به ترتیب برای توان های ۴ و ۸ مناسب میباشد.

در این مدل ها چون ما دچار **Flat** شده ایم در نمودارهای **learning curve** افزایش تعداد داده ها باعث کاهش خطا نمیشود.