

امتحان پایان ترم یادگیری ماشین

95104599

سید محمد حاجی میرسعیدی

گزارش:

هدف این پروژه پیشبینی مقدار آلاینده های هوا با استفاده از سرعت باد، دما، رطوبت، و همینطور مقدار خود آلاینده های هوا در ساعات قبل. دیتا این مسئله مربوط به شهر بسل است در کشور سوئیس و داده ها به صورت میانگین ساعتی در 6 ماه اخیر هستند که به طور کلی شامل 4391 داده میشود. قرار است مقدار دوز 6 آلاینده $NO_2, SO_2, O_3, PM, NO, CBC$ در این پروژ [پیدا شود. برای هر آلاینده یک مدل جداگانه استفاده شده دلیل آن در ادامه گفته شده است.

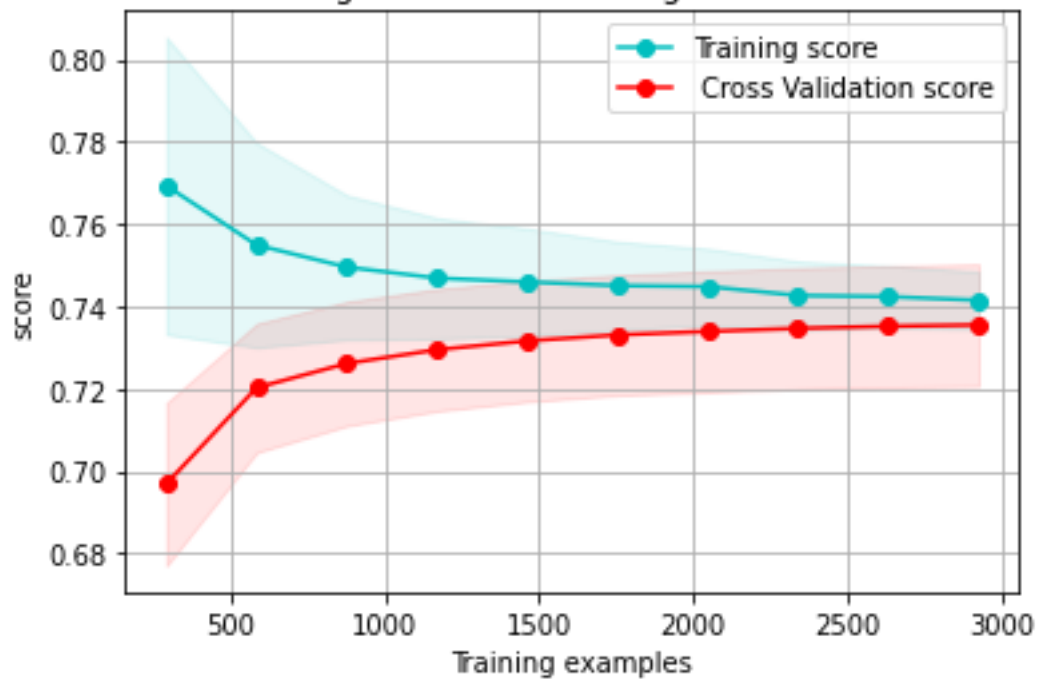
با توجه به اینکه مقدار آلاینده ها در زمان حاضر بسیار وابسته است به مقدار آلاینده ها در ساعات گذشته برای همین من مقدار 24 ساعت گذشته هر آلاینده را به عنوان 24 فیچر هر سَمپل به آن اختصاص دادم و این دلیل این بود که مجبور شدم مقدار هر آلاینده را جداگانه پیش بینی کنم زیرا مقدار گذشته آلاینده از مقدار الان آلاینده دیگر مستقل است. عامل دیگری که در مقدار آلاینده ها تاثیر دارد سرعت باد استو چون سرعت باد در ساعات گذشته مقدار آلاینده هوا در حال حاضر را مشخص میکند(در حقیقت تاثیر باد با تاخیر عمل میکند) من مقدار باد در 5 ساعت گذشته هر داده را نیز به عنوان 5 فیچر به مسئله دادم. مقدار دیگر تاثیر گزار مقدار رطوبت هوا بود که میشود به عنوان شاخصی از بارندگی روی آن تکیه کرد که مانند دو عامل قبلی دچار تاخیر است پس تا 3 ساعت قبل به عنوان فیچر انتخاب شده است. عمل دیگر دما است ولی چون دما کند تغییر است پس فقط در لحظه حاضر را به عنوان فیچر به ورودی میدهیم. عامل دیگر ساعت شبانه روز است که به صورت وان هات کدینگ به عنوان فیچر به مسئله میدهیم. دلیل اینکه از داده های قبل برای پیشبینی حال استفاده کردیم این بود که وابستگی زمانی داده از نظر فیزیکی زیاد به نظر میرسد و داده ها دارای کورلیشن زمانی بودند. یکی از روش هایی که روی مسئله تست کردم knn بود ولی متأسفانه فاصله خطای تست و ترین زیاد میشد برای همین از روش لینیئر رگرشن استفاده کردم.

محل تهیه دیتا ها همان سایتی است که در گزارش اول امتحان گفته شد.

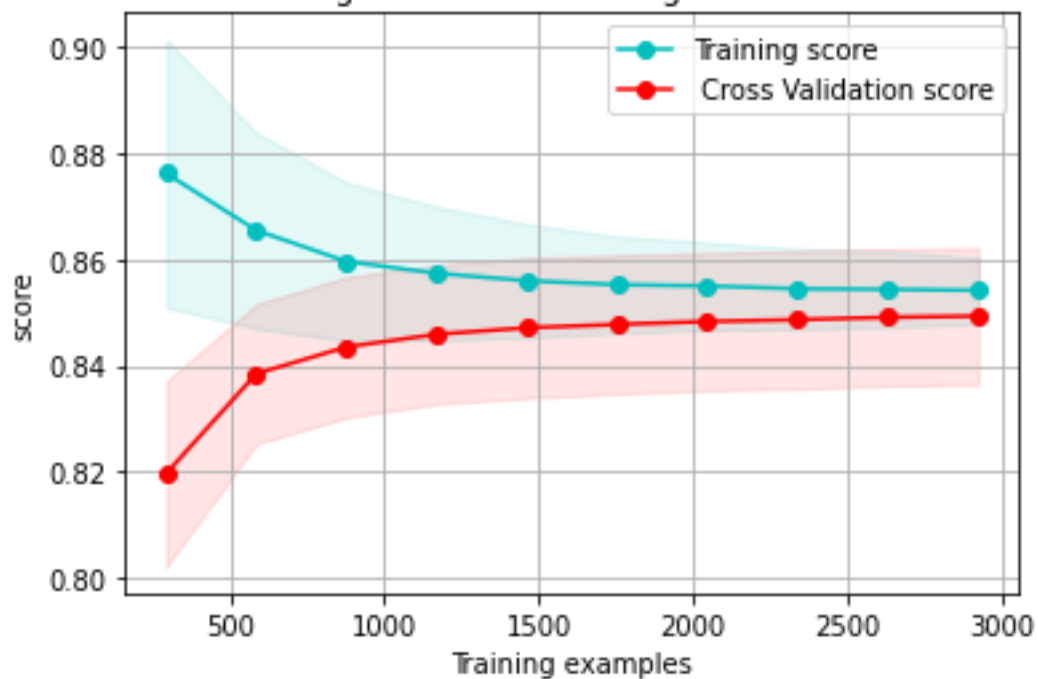
روشی که برای پیش بینی کردن استفاده شد لینیئر رگرشن بود چون عوامل بیان شده به ظاهر رابطه خطی خیلی زیادی دارند. نمودار های لرنینگ کرو به صورت زیر اند:

	O_3 [$\mu g/m^3$]	NO_2 [$\mu g/m^3$]	SO_2 [$\mu g/m^3$]	$PM_{2.5}$ [$\mu g/m^3$]	CPC [$1/cm^3$]	NO_x [$\mu g/m^3$ eq. NO_2]
0	3.7	31.1	0.9	13.5	7003.0	36.3
1	2.4	32.7	1.1	14.1	7322.2	40
2	1.9	32.2	1.2	15.9	7934.7	40.2
3	1.9	31.6	1.5	16.6	8642.7	41.6
4	2	31.5	1.2	16.6	7249.1	43.6
...
4386	161.2	6.8	0.7	5.8	7991.2	7.1
4387	144.4	8.5	0.9	5.1	7637.6	8.7
4388	123.1	9.7	0.9	5.5	8948.7	10
4389	100.1	10.5	0.6	7.7	8877.3	10.9
4390	104.3	7.4	0.5	7.4	7101.7	7.7

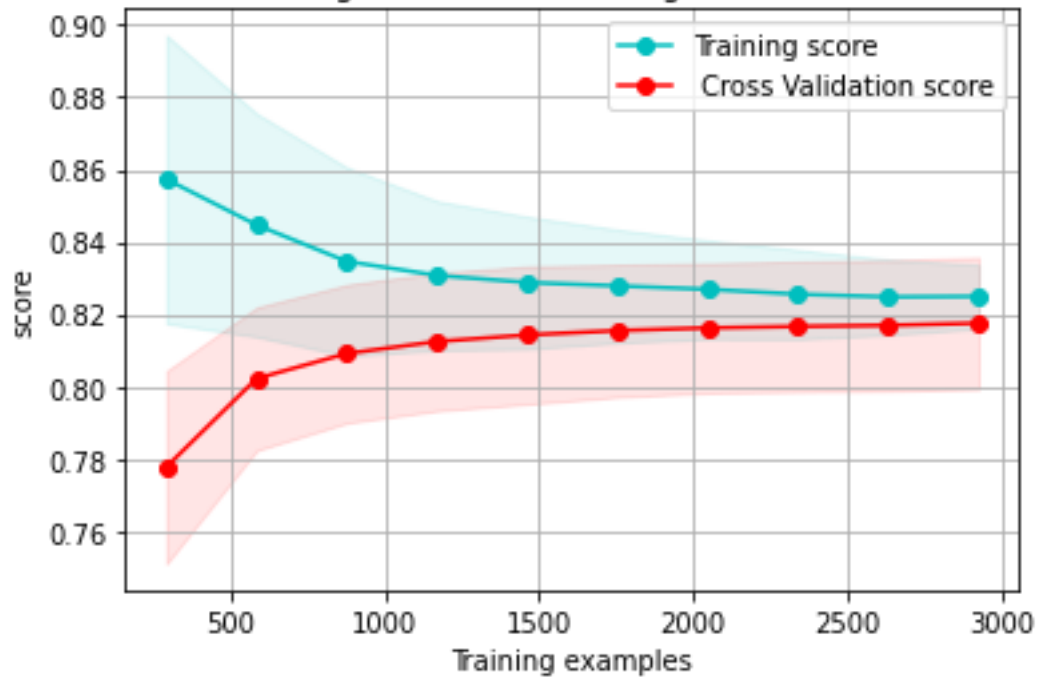
Learning curve for linear Regression for CBC



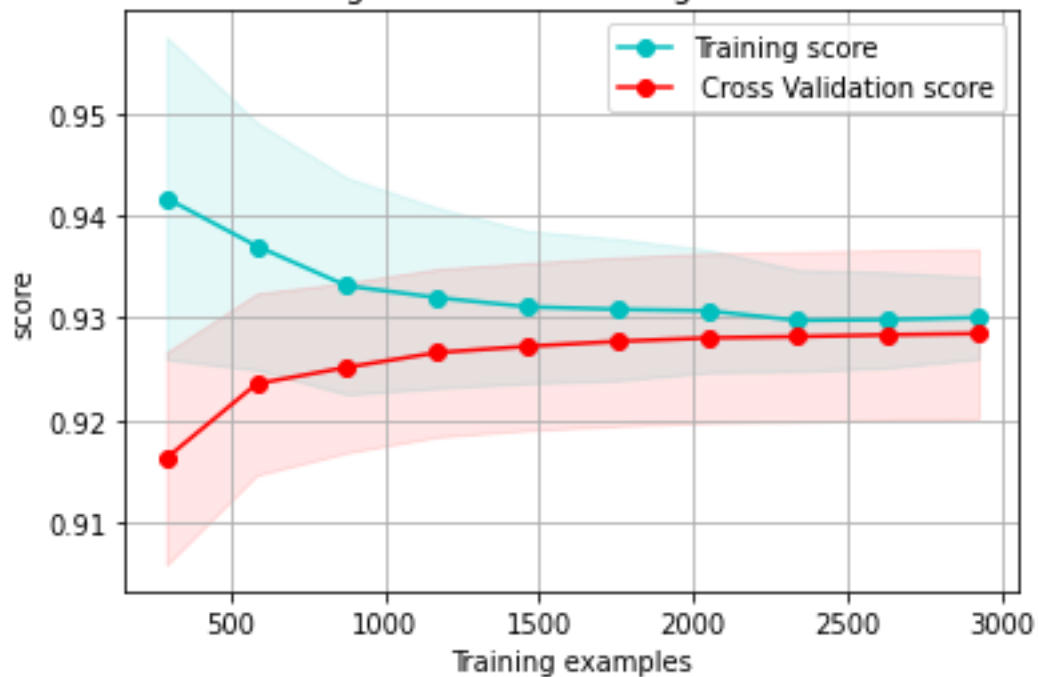
Learning curve for linear Regression for NO2



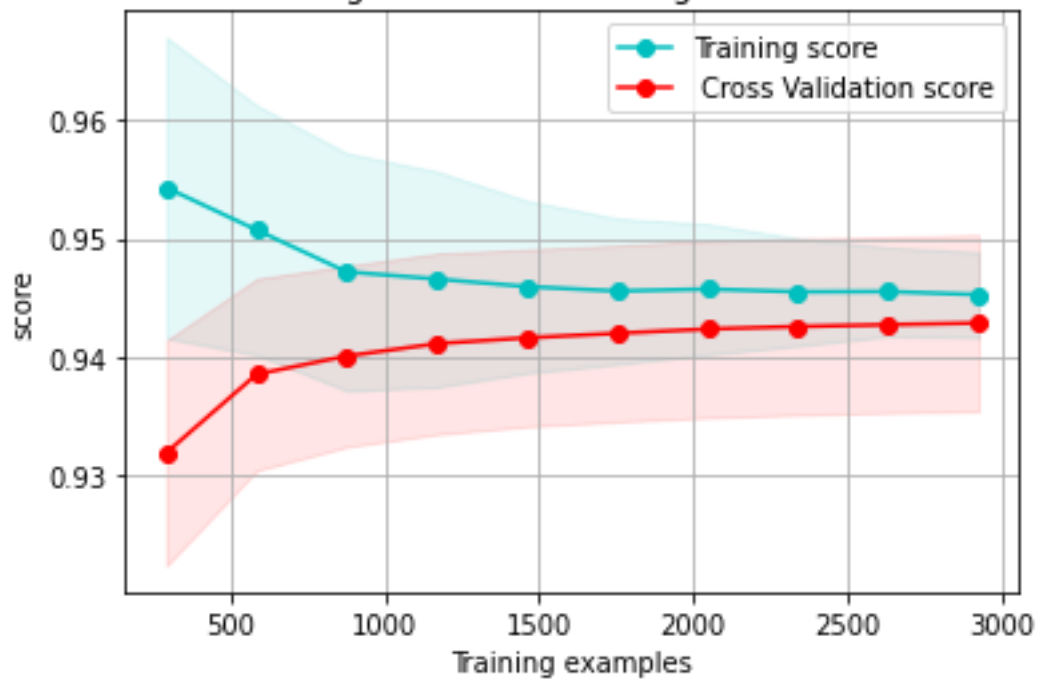
Learning curve for linear Regression for NOX



Learning curve for linear Regression for O3



Learning curve for linear Regression for PM



Learning curve for linear Regression for SO2

