به نام خدا



دانشگاه صنعتی امیرکبیر

(پلی‌تکنیک تهران)

درس یادگیری ماشین

استاد ناظرفرد

تمرین اول

علیرضا مازوچی

400131075

بخش اول: پرسش‌های تشریحی

سوال ۱

الف) نادرست؛ چنین گزاره‌ای همواره برقرار نیست. به عنوان مثال اگر مدل دچار بیش‌برازش شده باشد و باز پیچیدگی مدل را زیاد کنیم، خطای مدل در هنگام آموزش کم‌‌تر هم می‌شود ولی در موقع تست ممکن است خطا بیشتر شود. برای درک بهتر می‌توانید به تصویر سوال ۲ مراجعه کنید.

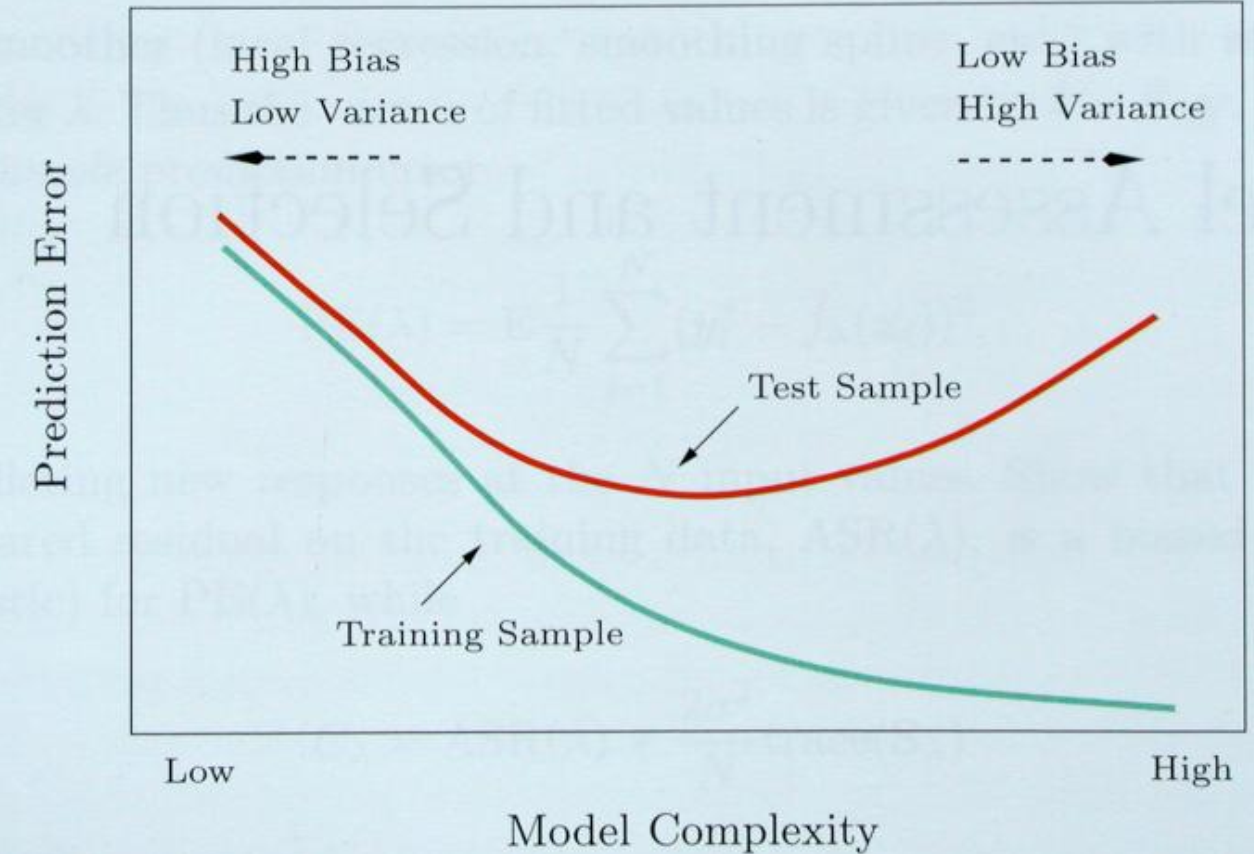
ب) درست اگر پیچیدگی مدل را ثابت فرض کنیم؛ پیچیدگی مدل و پارامتر‌های آن باید متناسب با تعداد داده و شرایط مسئله باشند. طبیعی است که در این شرایط و با فرض ثابت ماندن پیچیدگی مدل اگر داده‌های آموزش کمتر شود،‌ استعداد مدل برای بیش‌برازش زیادتر می‌شود. اگر هم بتوان مدل را ساده‌تر کرد که این مشکل جلوگیری است.

ج) نادرست؛ الزاما خیر! به عنوان مثال اگر مدل دچار کم‌برازش باشد، ‌پیچیده‌تر کردن به اندازه مدل باعث کاهش هم خطای آموزش و هم خطای تست خواهد شد. برای درک بهتر می‌توانید به تصویر سوال ۲ مراجعه کنید.

د) درست؛ به طور کلی معیار MAE در برابر نویز مقاوم‌تر است. داده‌های نویز داده‌هایی هستند که از الگوی واقعی داده‌های مسئله فاصله زیادی دارند. در معیار MSE‌ چون از توان دو برای فاصله استفاده می‌کند، مدل را برای داده‌های با فاصله زیاد از جمله داده‌های نویز بیشتر جریمه می‌کند درحالی که در MAE این‌چنین نیست. در معیار RMSE هم اگرچه از یک رادیکال استفاده می‌شود، اما در زیر رادیکال باز داده‌های نویز می‌توانند بخش مهم‌تری از خطا را تشکیل دهند.

سوال ۲

این اتفاق زمانی رخ می‌دهد که مدل مانند تصویر زیر دچار بیش‌برازش شده باشد. به این علت که مدل در زمان آموزش به نوعی داده‌ها را به جای یادگیری، حفظ کرده است و روی آن می‌تواند خطای پایینی داشته باشد. اما در هنگام اعتبارسنجی (یا تست) و با داده‌های جدید مدل می‌تواند خطای بالایی داشته باشد چون الگویی که توسط مدل یادگرفته شده است یک الگوی شدیدا وابسته به نمونه‌های آموزش است.



برای حل این مشکل می‌توان راهکار‌های زیر را پیشنهاد داد:

1. کاهش پیچیدگی مدل: منطقی‌ترین و متداول‌ترین راه این است که پیچیدگی مدل را کاهش دهیم تا بیش‌برازش از میان برود.
2. افزایش نمونه‌های آموزش: اگر نخواهیم پیچیدگی مدل را کاهش دهیم، می‌توان حجم داده‌های آموزش را افزایش داد تا تناسب این دو برقرار شود.
3. افزودن جمله منظم‌ساز: می‌توان به تابع خطای مسئله یک جمله منظم‌ساز اضافه کرد تا در هنگام آموزش مدل سعی کند با حداقل پیچیدگی به دقت مناسب برسد.

سوال ۳

سوال ۴[[1]](#footnote-1)[[2]](#footnote-2)

الف) الگوریتم معادله نرمال مناسب نیست زیرا به تعداد دفعه بالا باید برای هر داده ماتریس وارون را محاسبه کرد. دو روش Stochastic GD و Batch GD‌ به دلیل پردازش دسته‌ای مشکل زمان را ندارند. از نظر حافظه هم مشکلی ندارند چراکه کافی است تا تنها بخشی از داده در حافظه باشد. روش Batch GD هم مشکل زمان را ندارند ولی ممکن است از نظر حافظه‌ای مشکل‌ساز شود. اگر حافظه به اندازه کافی داشته باشیم این روش را هم می‌توان پذیرفت درغیر این صورت خیر.

ب) زمانی که از الگوریتم مهادله نرمال استفاده می‌شود مقیاس‌های متفاوت ویژگی‌ها موثر نیست و مشکلی ایجاد نمی‌کند اما برای الگریتم‌های مبتنی بر گرادیان نزولی این مسئله تاثیر منفی می‌گذارد. در backpropagation و برای ویژگی‌ها با مقیاس بزرگ نیاز است که از گام‌های بلند استفاده شود اما برای سایر ویژگی‌ها باید از گام کوچک‌تر استفاده کرد. در نتیجه یا دقت مناسب را به سبب لطمه خوردن به ویژگی‌ها با مقیاس کم آسیب خواهد دید و یا آنکه مجبور به صرف زمان زیاد خواهیم بود تا تمام ویژگی‌ها به مقادیر مناسب خود برسند. برای حل این مشکل می‌توان از روش Feature Scaling‌ استفاده کرد. با روش Feature Scaling تمامی ویژگی‌ها در یک مقیاس تقریبا یکسان قرار خواهند گرفت و سرعت همگرایی دسته الگوریتم‌های مبتنی بر گرادیان نزولی تسریع پیدا می‌کند. برای Feature Scaling روش‌های متعددی هست مثلا می‌توان از روش نرمال‌سازی min-max با فرمول زیر بهره جست. در این رابطهنرمال‌شده ویژگی خواهد بود.

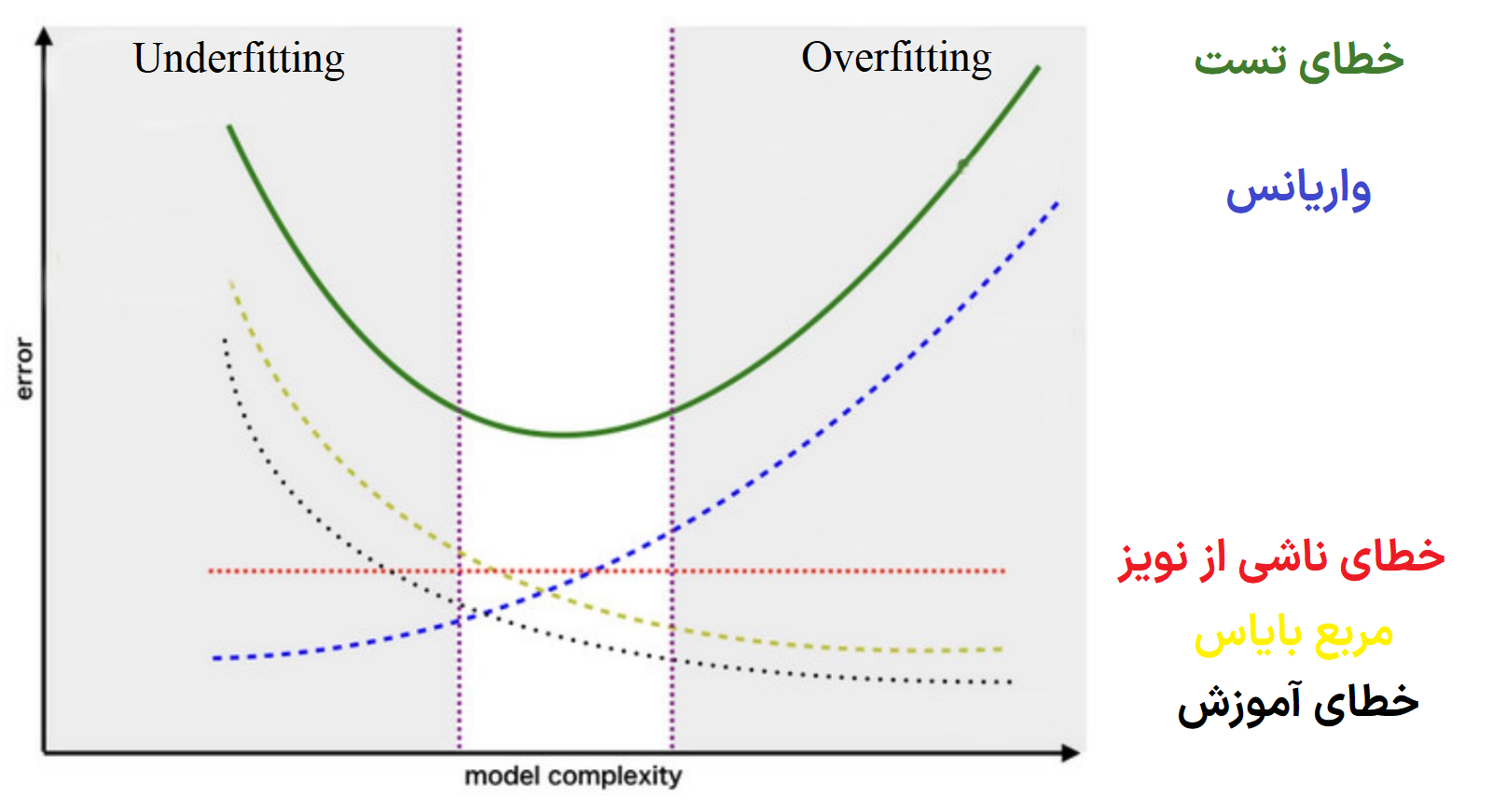
پس از اعمال نرمال‌سازی مذکور مقدار تمام ویژگی‌ها در رنج ۰ تا ۱ قرار خواهد گرفت.

سوال ۵

به طور کلی، با افزودن داده‌های جدید به مدلی که واریانس بالایی دارد، مشکلات ناشی از واریانس بالا کمتر می‌شود. قبل از افزودن داده‌های جدید مدل توجه زیادی به نمونه‌های موجود در مجموعه آموزشی دارد و در مواردی به جای توجه به الگوی‌کلی درگیر نمونه‌ها می‌شود. پس از افزودن داده‌های بیشتر، مدل بیشتر کنترل می‌شود و یک داده نویز به تنهایی نمی‌تواند مانند سابق در نظر مدل اهمیت داشته باشد.

به طور کلی، با افزودن داده‌های جدید به مدلی که بایاس بالایی دارد، تاثیر چندان مثبتی برای مدل رخ نخواهد داد. توجه کنید که پیش از افزودن داده به مدل،‌ مدل اهمیت کافی به داده‌های موجود نمی‌دهد و الگوی داده‌ها را ساده‌تر از چیزی که باید باشد تصور می‌کند. بدیهی است که در این ظرایط افزودن داده‌های بیشتر در چیزی که مدل یادگرفته است تاثیر جدی‌ای نخواهد داشت و مدل همچنان الگوی سابق را پیشنهاد می‌دهد. در این شرایط نه تنها مشکل بایاس حل نشده است بلکه بیشتر هم شده است!

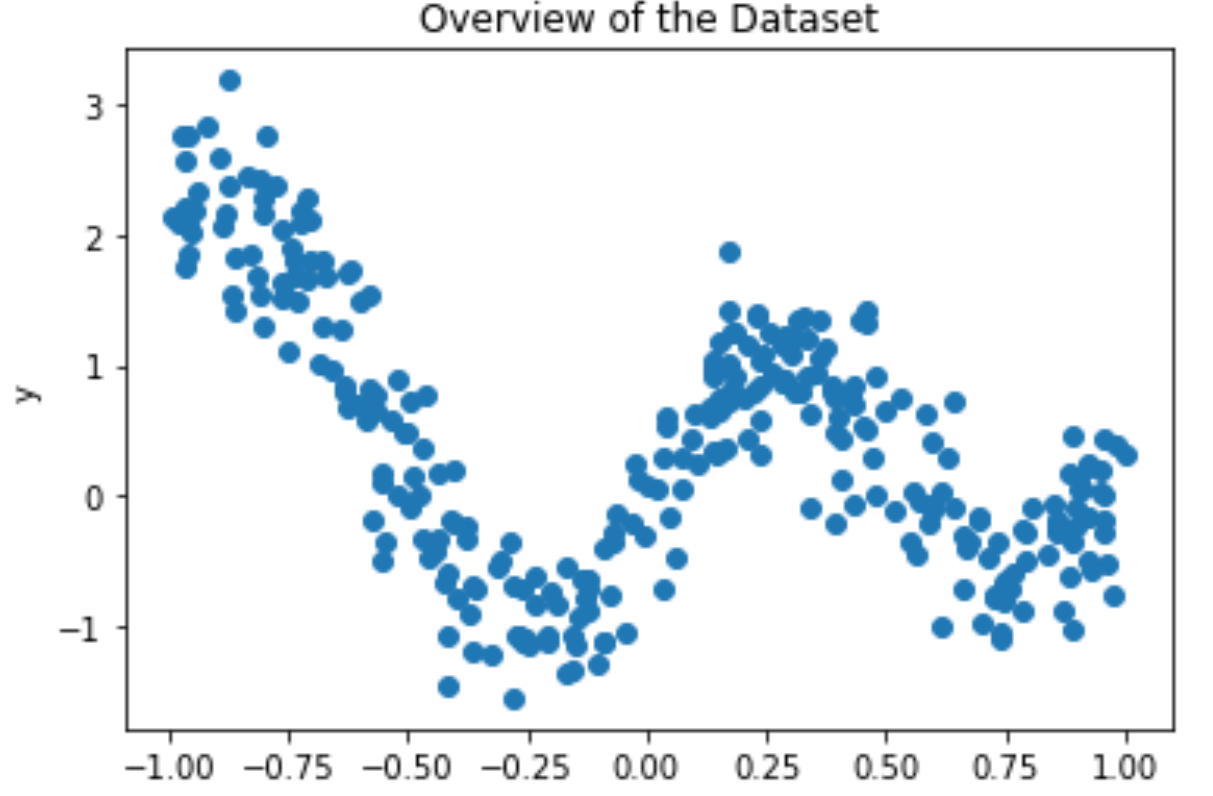
سوال ۶



بخش دوم: پیاده‌سازی

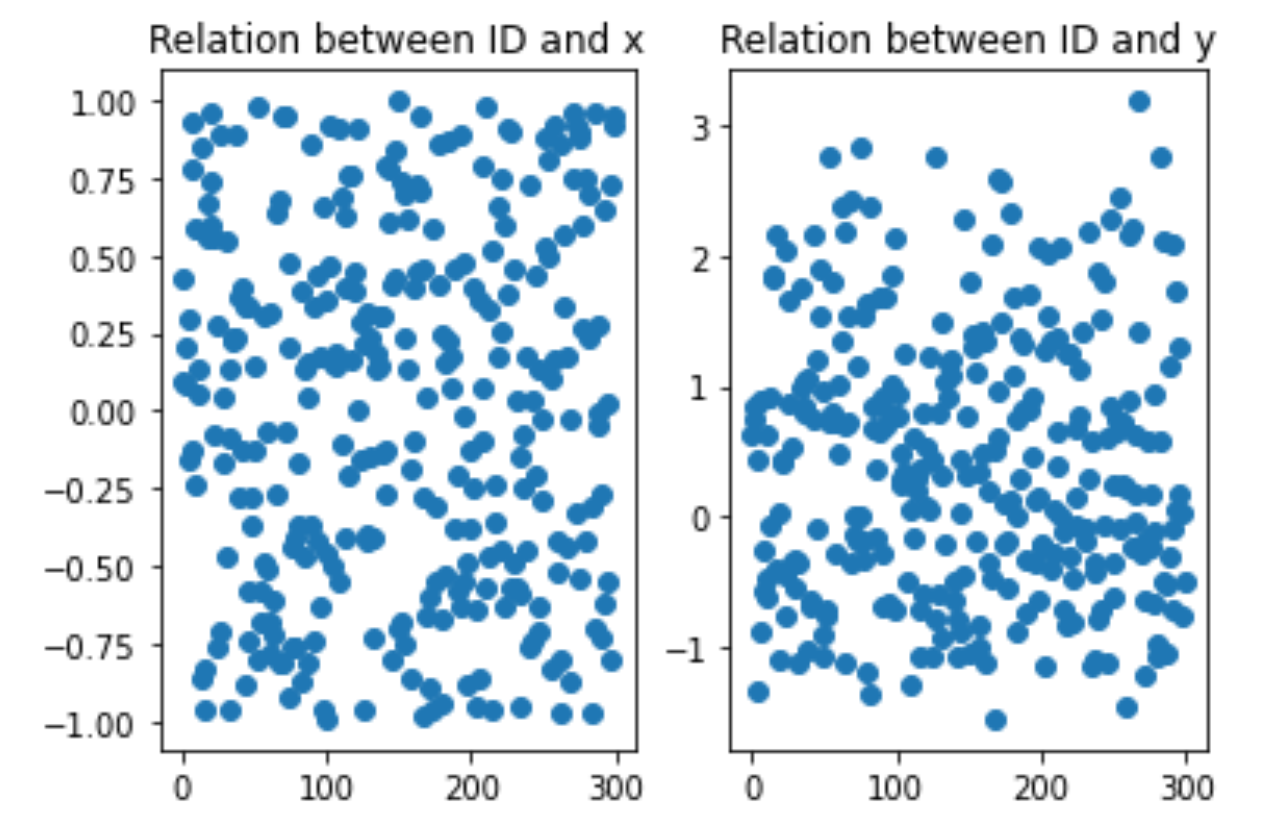
سوال ۱

الف)

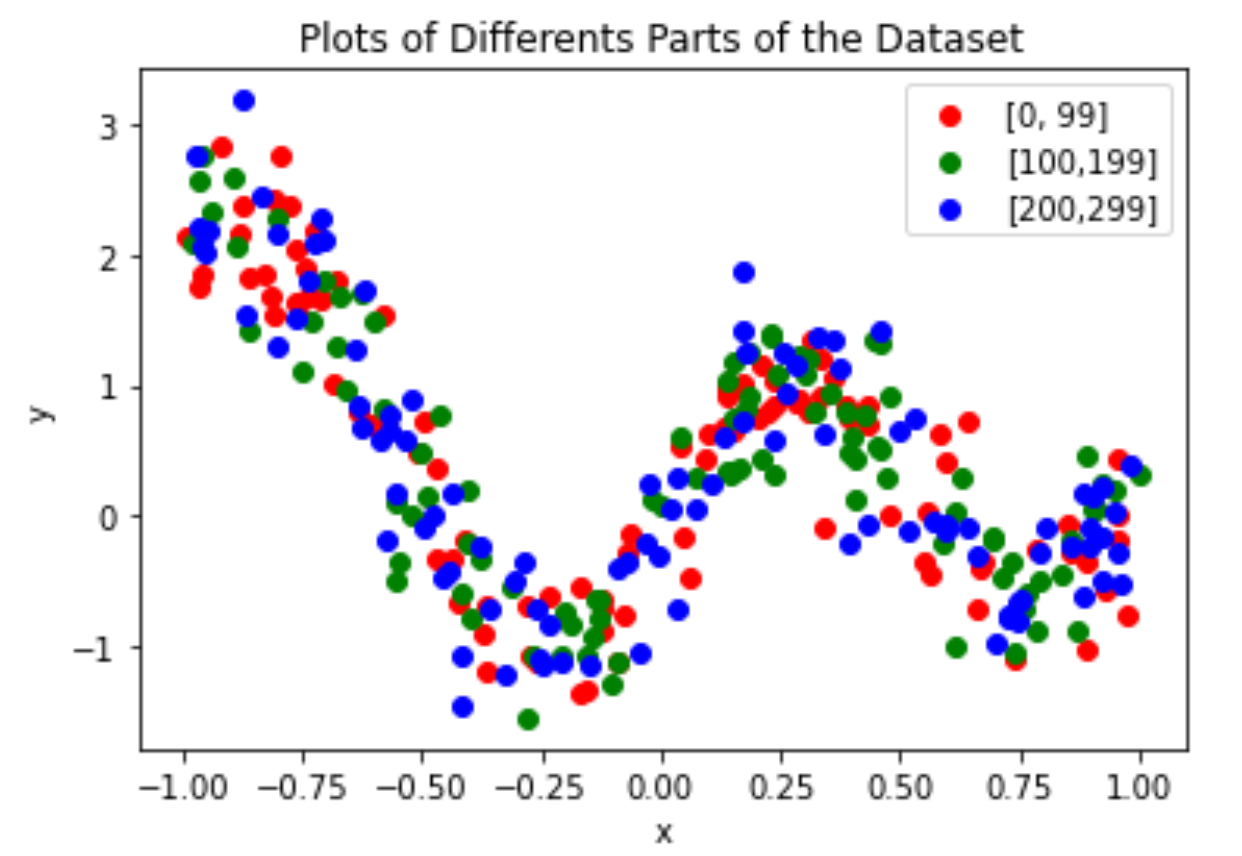


ب) در برخی از مجموعه‌های داده ممکن است ترتیب اولیه داده‌ها بامعنی باشد. مثلا مقدار یکی از ویژگی‌ها به صورت افزایشی مرتب شده باشد یا داده‌های مربوط به یک کلاس در ابتدا بیاید و داده‌های کلاس دیگر در انتها! چنین چیزی می‌تواند مشکلاتی را به وجود بیاورد. مثلا موقع تقسیم داده‌ها به مجموعه آموزشی و تست، چون عمدتا تکه اول داده‌ها برای آموزش و تکه دوم برای تست برداشته می‌شوند،‌ دو مجموعه از همدیگر متفاوت خواهند شد و مدل در زمان تست در شرایط جدیدی قرار خواهد گرفت. همچنین در هنگام آموزش هم اگر قرار باشد داده‌ها به صورت دسته‌ای به مدل داده شود بهتر است مجموعه آموزشی شافل شده باشد تا هر دسته شرایط نسبتا مشابهی با کل داده‌ها داشته باشد.

برای بررسی اینکه آیا نیاز به شافل کردن وجود دارد یا خیر من دو آزمایش را انجام داده‌ام. ابتدا بررسی کرده‌ام که آیا مقدار یک ویژگی با افزایش شماره آیدی دارای الگوی خاصی است یا نه؟ برای دو ویژگی x و y نمودار‌های زیر حاصل شد که نشان می‌دهد ارتباطی میان آیدی و این دو ویژگی وجود ندارد:



نهایتا بررسی کردم که آیا قسمت‌های مختلف مجموعه‌داده الگویی مشابه هم دارند یا خیر؟ برای این کار مجموعه‌داده را به سه قسمت تقسیم کردم و آن‌ها را ترسیم کردم. به نظر می‌رسد الگوی هر سه قسمت مشابه هم دیگر است و مجموعه‌داده نیازی به شافل شدن ندارد.



1. https://gist.github.com/byelipk/345ee92c42f579a9dc1938b0bb86be2e [↑](#footnote-ref-1)
2. https://www.atoti.io/when-to-perform-a-feature-scaling/ [↑](#footnote-ref-2)