

به نام خدا



دانشگاه صنعتی خواجه نصیرالدین طوسی دانشکده برق

شبیه سازی و مدلسازی

گزارش تمرین شماره 2

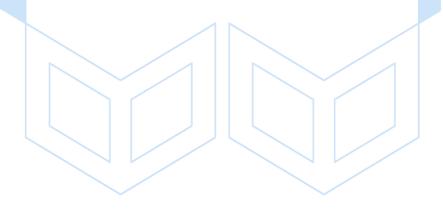
سجاد رجبی باغستان 40005393

استاد : جناب دکتر مهدی علیاری

بهمن 1402

فهرست مطالب

شماره صفحه		عنوان
3	: چکیده	بخش ۱
4	: سوالات تحليلي	بخش ۱
4	ى اول	سوال
5	: سوالات شبیه سازی	بخش۲
5	ے اول	سوال
	ي دوم	
7	ى سوم	سوال
11		مراجع.



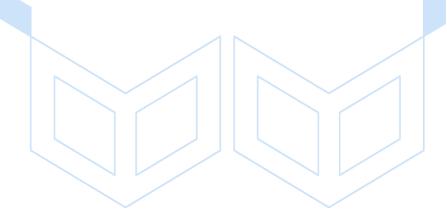
چکیده

در این تمرین، به بررسی و مقایسه دو روش انتخاب ویژگی Forward Selection و backward elimination

Forward Selection یک روش انتخاب ویژگیهای مهم در مدلهای رگرسیون است. در این روش، از یک مدل اولیه شروع میشود که فقط یک ویژگی دارد. سپس ویژگیهای دیگر به تدریج اضافه میشوند و هر بار اثر اضافه شدن یک ویژگی را ارزیابی میکنیم. ویژگیای که اضافه شدن آن باعث بهبود بیشتری در عملکرد مدل میشود، به مدل اضافه میشود.

Backward Elimination یک روش انتخاب ویژگی در مدلهای رگرسیون است که به تدریج ویژگیهای غیرضروری را از مدل حذف میکند تا به بهترین مدل برسد. در این روش، از یک مدل اولیه که شامل تمام ویژگیها است شروع میشود، سپس ویژگیهای غیرضروری یکییکی حذف میشوند و عملکرد مدل با هر حذف ارزیابی میشود.

مزایای این روش شامل کاهش ابعاد دادهها، سرعت بالا در مواقعی که تعداد ویژگیها زیاد است، و قابلیت استفاده از معیارهای مختلف برای ارزیابی مدل است. از معایب آن میتوان به این اشاره کرد که ممکن است در مواقعی ویژگیهای مهم حذف شوند و همچنین این روش بیشتر در مواقعی مؤثر است که تأثیرات ویژگیها به طور مستقل از هم نباشند.



بخش ١: سوالات تحليلي

سوال اول

در بحث مدل سازی ،هدف این است که مدلی ارائه شود که نه بیش از حد بی دقت باشد و نه بیش از حد بی دقت باشد و نه بیش از حد با دقت (اصطلاحا overfit یا underfit نباشد). با داشتن رگرسور های فراوان برای مدل سازی ،مدل ارائه شده به صورت خیلی دقیق عمل میکند به طوری که مدل ما overfit میشود ، همچنین با داشتن تعداد زیادی رگرسور ، مقدار زیادی نویز وارد مدل ما میشود زیرا هر رگرسور همراه با نویز است.

در ضمن این موضوع را باید در نظر داشت که ممکن است اثر یک رگرسور نسبت به بقیه رگرسور ها خیلی بیشتر باشد و همین امر باعث این میشود که استفاده از بقیه رگرسور ها ناصرفه باشد.

دلایل ضرورت subsection:

- به طور کلی در واقعیت تعداد رگرسور ها میتواند بسیار زیاد باشد و همین امر باعث پیچیدگی زیاد مدل از نظر محاسباتی شود به طوری که انجام این محاسبات ناممکن و یا بسیار زمان بر است.
- در بسیار مواقع سیستم تنها از تعداد محدودی رگرسور تاثیر اصلی را میپذیرد و بقیه رگرسور
 ها با توجه به هزینه ای که برای مدل سازی صرف میشود (زمان، محاسبات، تجهیزات،...)غیر ضروری هستند.

فواید subsection:

- کاهش زمان و منابع
- بهتر شدن مدل و دور شدن مدل از overfitting

بخش۲: سوالات شبیه سازی

سوال اول

در این سوال بعد از import کردن دیتا ها ، رگرسور ها را یک به یک انتخاب کرده و اضافه میکنیم این کار را تازمانی ادامه میدهیم که کمترین تغییرات را برای ساخت y پیدا کنیم. رگرسور هایی که انتخاب میشوند بیشترین کواریانس را با y دارند.

Result:

Selected Regressors:

1 2 3

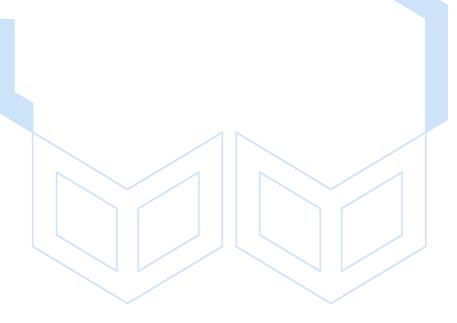
Parameters:

0.1719

0.2174

0.3016 0.2892

>>



سوال دوم

در این سوال بعد از import کردن دیتا ها ، رگرسور ها را یک به یک انتخاب کرده و حذف میکنیم این کار را تازمانی ادامه میدهیم که بیشترین تغییرات را در y مشاهده کنیم.

Result:

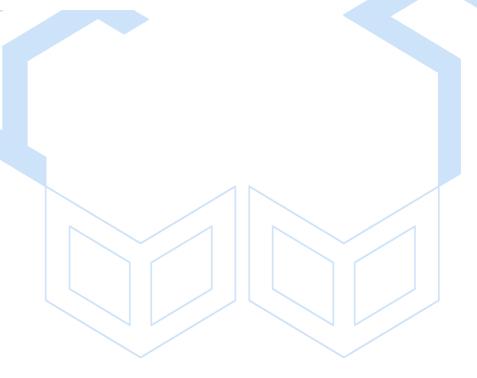
Selected Regressors:

1 2 3 4

Parameters:

0.1719 0.2174 0.3016 0.2892

>>



سوال سوم

جواب نهایی در هردو روش یکسان است .

تفاوت در سرعت پاسخگویی:

نکته ای که این دو روش را از هم متمایز میکند الگوریتم آن است که باعث تغییر در سرعت اجرای کد و انجام محاسبات میشود .

Forward Selection: Elapsed time is 0.012016 seconds.

Backward Elimination: Elapsed time is 0.005213 seconds.

با توجه به مقادیر به دست آمده در رابطه با زمان مورد نیاز برای انجام محاسبات نتیجه گیری میشود که زمان اجرا روش Forward Selection تقریبا دو برابر Backward Elimination است و این موضوع زمانی بسیار مهم می شود که حجم دیتا ستی که در اختیار ما است بسیار زیاد باشد ،دراین صورت احتمالا ترجیح بر این است که از روش کوتاه تر استفاده شود .

تفاوت در سرعت اجرا بین روشهای Backward Elimination و Forward Selection ممکن است به دلایل زیر باشد:

- انتخاب ویژگیها: در روش Forward Selection، ویژگیهایی که به مدل اضافه میشوند، بر اساس بهبودی که در عملکرد مدل به دنبال اضافه شدن هر ویژگی حاصل میشود، انتخاب میشوند. این میتواند منجر به انتخاب سریعتر ویژگیهای مهمتر و کارا باشد.
- حذف ویژگیها: در روش Backward Elimination، ما به تدریج ویژگیهای غیرضروری را از مدل حذف می کنیم. این عملیات حذف ممکن است در مواردی زمانبرتر باشد، مخصوصاً اگر تعداد ویژگیها زیاد باشد.

در کل، اگرچه هر دو روش می توانند به بهبود عملکرد مدل منجر شوند، اما عملکرد و سرعت آنها ممکن است به ویژگیهای داده، اندازه داده و پیچیدگی مدل بستگی داشته باشد.

تفاوت در صحت پاسخگویی:

هرچند که در این آزمایش مشخص نشد اما با توجه به منابع موجود و تجربه های دیگر نتایج این دو روش در گاهی اوقات متفاوت به دست می آید به این صورت که روش Forward Selection با خطا کمتر و صحت بیشتری نسبت به روش Backward Elimination عمل میکند.

احتمال خطا در روش Backward Elimination نسبت به Backward جمکن است به دلایل زیر بیشتر باشد:

- شناسایی تأثیرات متقاطع: در بعضی موارد، ویژگیها ممکن است با هم تعامل داشته باشند و حذف یک ویژگی می تواند تأثیرات متقاطع دیگر ویژگیها را تغییر دهد. این تعاملات ممکن است توسط روش Backward Elimination در نظر گرفته نشوند که باعث ایجاد خطاهای نادرست در انتخاب و حذف ویژگیها می شود.
- انتخاب متغیرهای نادرست: در صورتی که یک ویژگی به طور نادرست از مدل حذف شود، ممکن است اطلاعات مهمی را که مدل باید به طور صحیح استفاده کند، از دست بدهیم. این موضوع میتواند به افزایش خطا در پیشبینیهای مدل منجر شود.
- استقرار متغیرهای تصادفی: در صورتی که متغیرهایی به طور تصادفی از مدل حذف شوند بدون
 در نظر گرفتن اثرات واقعی آنها، این می تواند به خطاهای نادرست در مدل منجر شود.

به طور کلی، عدم در نظر گرفتن اثرات متقاطع و تداخلات بین متغیرها و انجام حذفها به طور نادرست ممکن است باعث افزایش احتمال خطا در روش Backward Elimination شود. از طرفی، روش Selection به طور تدریجی ویژگیهای مهم را به مدل اضافه می کند و این می تواند باعث کاهش احتمال خطا در انتخاب متغیرها شود.

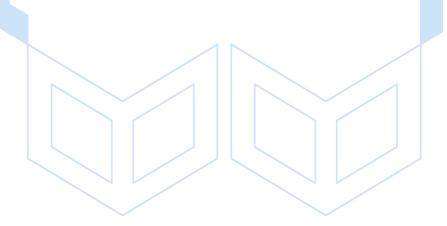
تفاوت در الگوریتم:

:Forward Selection

- 1- شروع با ویژگی خالی: در این روش، مدل با یک ویژگی شروع میشود و سپس ویژگیهای دیگر به ترتیب اضافه میشوند تا به بهترین مدل برسیم.
- 2 افزودن ویژگیها: ویژگیهایی که با اضافه شدن به مدل، منجر به بهبود عملکرد مدل میشوند، به مدل اضافه میشوند.

:Backward Elimination

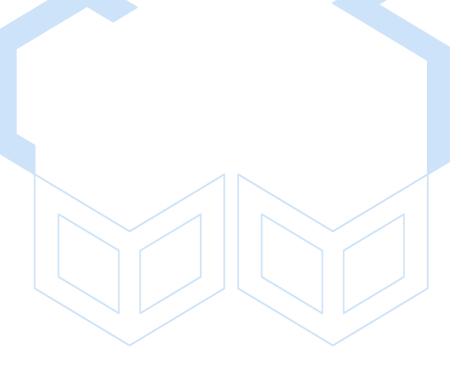
- 1- شروع با تمام ویژگیها: در این روش، مدل با تمام ویژگیها شروع میشود و سپس ویژگیهای غیرضروری به تدریج حذف میشوند.
- 2- حذف ویژگیها: ویژگیهایی که موجب کاهش خطای مدل نمیشوند، یکییکی از مدل حذفمیشوند تا به بهترین مدل برسیم.



خلاصه نتايج:

Forward Selection	Backward Elimination	ویژگی
شروع با یک ویژگی	شروع با تمام ویژگیها	انتخاب اوليه رگرسور
اضافه کردن ویژگیهایی که به عملکرد مدل کمک میکنند	حذف ویژگیهایی که عملکرد را بهبود نمیدهند	افزایش ویژگیها
کمتر	بيشتر	تعداد مراحل
ممكن است بيشتر باشد	ممكن است كمتر باشد	زمان اجرا
كمتر	بيشتر	احتمال خطا
بيشتر	كمتر	حساسیت به تعداد ویژگیها

جدول 1 : مقايسه backward elimination و



مراجع

Websites:

https://chat.openai.com/

https://en.wikipedia.org

https://www.sciencedirect.com/topics/computer-science

https://www.simplilearn.com

book:

Nonlinear system identification 2020 nelles

