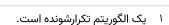
ياسخنامه

پاسخنامه کوئیز جلسه 08 دوره علم داده، یادگیری ماشین و هوش مصنوعی - فرزاد مینویی

🗸 ۴ کدام گزینه درباره داده گمشده نادرست است؟ ۱ حضور داده گمشده اندازه نمونه را کاهش میدهد. ۲ حضور داده گمشده ممکن است در نمونه سوگیری ایجاد کند. ۳ داده گمشده میتواند بعلت نحوه جمعآوری داده ایجاد شود. اگر کمتر از 10 درصد مشاهدات حاوی داده گمشده باشند، آنها را از نمونه حذف ン*۲− منظور از فرآیند Missing Completely at Random (MCAR) چیست؟ ۱ نمونهگیری از جامعه کاملاً تصادفی است. ۲ حضور داده گمشده در یک متغیر کاملاً تصادفی است. حضور داده گمشده در یک متغیر تا حدی براساس سایر متغیرها قابل پیشبینی است. ۴ همه موارد ❤ • کدام گزینه درباره روشهای Imputation در دادههای گمشده نادرست است؟ رویکرد Complete Case Approach از همه دادههای در دسترس استفاده ۲ رویکرد Mean/Median Substitution برای دادههای رستهای کاربرد ندارد. ۳ الگوریتم Micsing at Random (MAR) برای حالت Missing at Random (MAR) بسیار مناسب است. همه الگوریتمهای یادگیری ماشین با رویکرد Using All-Available Data سازگار

نیستند.

🗸 ۴ كدام گزينه درباره الگوريتم MICE نادرست است؟



برای برآورد داده گمشده از الگوریتمهای یادگیری ماشین مانند Light GBM استفاده میکند.

۳ از آن برای برآورد داده گمشده در متغیرهای عددی و رستهای استفاده میگردد.

۴ اندازه نمونه در این روش بسیار کاهش مییابد.

ک*۵− در دیتافریم زیر، مقادیر گمشده را با میانگین هر ستون جایگزین و ضریب همبستگی پیرسون بین x و y را برآورد کنید.

х	У
12.1	0.2
14	0.41
<na></na>	0.32
10	0.19
2	0.05
23	0.67
<na></na>	0.11
15	0.32
7	<na></na>
11	0.23
	12.1 14 <na> 10 2 23 <na></na></na>

٥.77 ١

0.23 ٢

0.95 ٣

✓ 0.86 °

*۶- در دیتافریم زیر، از رویکرد Complete Case Approach استفاده و ضریب همبستگی پیرسون بین x و y را برآورد کنید.

۶*	igoremsize

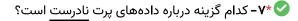
	x	у
0	12.1	0.2
1	14	0.41
2	<na></na>	0.32
3	10	0.19
4	2	0.05
5	23	0.67
6	<na></na>	0.11
7	15	0.32
8	7	<na></na>
9	11	0.23

0.//		0.77	١
------	--	------	---

0.23 ٢

✓ 0.95 m

۰.86 ۴



- ۱ داده پرت میتواند نماینده بودن نمونه از جامعه را زیر سوال ببرد.
- نتایج برخی از الگوریتمهای یادگیری ماشین میتواند به شدت تحت تاثیر داده پرت قرار گیرد.
 - ۳ داده پرت میتواند تحت تاثیر خطا در فرآیند جمعآوری داده ایجاد شود.
- ۴ یک مشاهده نمیتواند از منظر ترکیب چند متغیر مختلف داده پرت حساب گردد. 🗸

√*√منظور از Masking در تحلیل داده پرت چیست؟

روش شناسایی داده پرت خود تحت تاثیر داده پرت قرار گیرد.

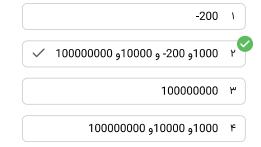
- ۲ داده پرت از منظر چند متغیر شناسایی شود.
- ۳ داده پرت تنها از منظر یک متغیر شناسایی شود.
- ۴ داده پرت بر روی الگوریتم یادگیری ماشین اثر بگذارد.

• ۲۰ و استفاده از متد کلاسیک، کدامیک از درایههای زیر بعنوان داده پرت شناسایی میشود؟ و ۲۰ میرود؟ ۱۰ میرود ۱ میرود ۱۰ میرود ۱ میرود ۱ میرود

100000000, 10000, 10000, 4, 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 200

	1000 و 200- و 10000 و 10000000 و	١
	100000000 و10000 و	۲
/	100000000	۳
	هیچکدام	۴

√*۰۰- با استفاده از متد توکی (Tukey's Method)، کدامیک از درایههای زیر بعنوان داده پرت شناسایی میشود؟



Porsline

با پُرسلاین فرم خود را طراحی و ارسال کنید و گزارشهای آماری دریافت نمایید.