

پاسخنامه

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

✓ ۱- 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 باعث نحوه جمع‌آوری داده ایجاد شود.
- ✓ ۴ اگر کمتر از 10 درصد مشاهدات حاوی داده گمشده باشند، آنها را از نمونه حذف می‌کنیم.

✓ 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.
- ۳ حضور داده گمشده در یک متغیر تا حدی براساس سایر متغیرها قابل پیش‌بینی است.
- ۴ همه موارد

✓ ۳- کدام گزینه درباره روش‌های Imputation در داده‌های گمشده نادرست است؟

- ✓ ۱ رویکرد Complete Case Approach از همه داده‌های در دسترس استفاده می‌کند.
- ۲ رویکرد Mean/Median Substitution برای داده‌های رسته‌ای کاربرد ندارد.
- ۳ الگوریتم MICE برای حالت Missing at Random (MAR) بسیار مناسب است.
- ۴ همه الگوریتم‌های یادگیری ماشین با رویکرد Using All-Available Data سازگار نیستند.

✓* ۴- کدام گزینه درباره الگوریتم MICE نادرست است؟

- ۱ یک الگوریتم تکرارشونده است.
- ۲ 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.

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

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

- ۱ 0.77
- ۲ 0.23
- ۳ 0.95
- ✓ ۴ 0.86

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

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

۱ 0.77

۲ 0.23

✓ ۳ 0.95

۴ 0.86

✓*۷- کدام گزینه درباره داده‌های پرت نادرست است؟

۱ داده پرت می‌تواند نماینده بودن نمونه از جامعه را زیر سوال ببرد.

۲ 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.

✓*۸- منظور از Masking در تحلیل داده پرت چیست؟

✓ ۱ 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.

✓ ۹- با استفاده از متد کلاسیک، کدامیک از درایه‌های زیر بعنوان داده پرت شناسایی می‌شود؟

100000000, 10000, 1000, 4, 4, 4, 4, 4, 3, 3, 3, 3, 3, 2, 2, 2, 2, -200

۱ 1000 و -200 و 10000 و 100000000

۲ 1000 و 10000 و 100000000

✓ ۳ 100000000

۴ هیچ‌کدام

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

100000000, 10000, 1000, 4, 4, 4, 4, 4, 4, 3, 3, 3, 3, 3, 2, 2, 2, 2, -200

۱ -200

✓ ۲ 1000 و -200 و 10000 و 100000000

۳ 100000000

۴ 1000 و 10000 و 100000000

Porsline

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