ДЕРЕКТЕРДІ ИНТЕЛЛЕКТУАЛДЫ ТАЛДАУ

Дәріс 7

NumPy: массивтер және сандық есептеулер

Оспан Ә.Ғ., PhD, доцент-зерттеуші

ҚазҰУ, Жасанды интеллект және Big Data кафедрасы

2026-2027 оқу жылы

Дәрістің мақсаты мен мазмұны

Мақсат және оқу нәтижелері

- shape, dtype және массив осьтерінің мағынасын түсіндіру.
- Тілімдерді, маскаларды және broadcasting қолдану.
- Сандық есептеулерді тексеріп, статистикаларды түсіндіру.

Дәріс мазмұны

1. **ndarray құрылымы**
2. **Осьтер және индекстеу**
3. **Көшірмелер мен маскалар**
4. **Пішіндер және broadcasting**
5. **Статистикалар**

ndarray: пішін, тип және ортақ операция

Пішін

- shape осьтердің ұзындықтарын береді.
- Вектор мен матрицаға әртүрлі индекстер қажет.

Тип

- dtype ұяшықтың бейнеленуін сипаттайды.
- Аралас деректер ортақ типті өзгертуі мүмкін.

Векторлау

- Операция тұтас массивке қолданылады.
- Қысқа формула нәтижені бақылауды жоймайды.

Пішін мен dtype — сандық есептеу келісімінің бөлігі.

Массив паспортын қалай оқу керек

Қасиет немесе өрнек	3x4 матрицасы үшін	Мағынасы
shape	(3, 4)	Осьтердің ұзындықтары
ndim / size	2 / 12	Осьтер / элементтер саны
dtype	Мысалы, float64	Бір ұяшықтың типі
a[1, 2]	Бір скаляр	Екінші жол, үшінші баған
a[:, 2]	Пішіні (3,)	Бағандар осі жойылды
a[:, 2:3]	Пішіні (3, 1)	Бағандар осі сақталды

Матрица құрып, қосындыларын тексереміз

Python мысалы

```
import numpy as np

grid = np.arange(1, 13).reshape(3, 4)
print(grid)
print(grid.sum(axis=0))
print(grid.sum(axis=1))
print(grid.shape, grid.ndim, grid.size)
```

Нәтиже: axis=0: [15, 18, 21, 24]; axis=1: [10, 26, 42].

Құру

- arange оң жақ шекті қамтымайды.
- reshape элементтер санын сақтайды.

Тексеру

- axis=0: әр бағанға бір нәтиже.
- axis=1: әр жолға бір нәтиже.

Ось жинақталу бағытын атайды

axis=0

- Жолдар осі жинақталады.
- (n, p) матрицасы (p,) векторын береді.
- Әр белгінің көрсеткішін аламыз.

axis=1

- Бағандар осі жинақталады.
- (n, p) матрицасы (n,) векторын береді.
- Әр нысанның көрсеткішін аламыз.

Алдымен осьтердің мағынасын атаңыз, содан кейін axis таңдап, кодты орындауға дейін нәтиже пішінін болжаңыз.

Жад көрінісі және тәуелсіз көшірме

Көрініс, view

- Негізгі тілім әдетте жадты бөліседі.
- Өзгеріс бастапқы нұсқаға әсер етуі мүмкін.
- Жаңа атау жаңа деректерді білдірмейді.

Көшірме, copy

- Тәуелсіз мәндерді сақтайды.
- Өзгертілетін тәжірибеге жарамды.
- Қосымша жад талап етеді.

Өзгеріске дейін қай нысан сақталуы тиіс екенін шешіп, көшірмені алдын ала жасаңыз.

Тілімді өзгертудің көрінетін салдары

Python мысалы

```
base = np.arange(1, 10).reshape(3, 3)
view = base[:2, :2]
view[0, 0] = -1
independent = base[:2, :2].copy()
independent[0, 0] = 100
print(base[0, 0], independent[0, 0])
print(np.shares_memory(base, view))
```

Нәтиже: -1 100; shares_memory(base, view) True қайтарады.

Ортақ жад

- view өзгерісі base массивін өзгертеді.
- Тілім деректерді көшіруді үнемдейді.

Тәуелсіздік

- copy келесі меншіктеуді оқшаулайды.
- Бірнеше белгілі ұяшықты тексереміз.

Маска: ұяшықтар не тұтас жолдар таңдау

1

Таңдау нысанын атау

Жеке ұяшықтарды, жолдарды немесе бағандарды таңдаймыз.

2

Шарт құру

Салыстыру bool береді; шарттар & немесе | арқылы біріктіріледі.

3

Пішінді сәйкестендіру

Маска ұзындығы таңдалған оське сәйкес келеді.

4

Нәтижені тексеру

shape, таңдалған орындар саны және белгілі мысал қаралады.

- grid[grid > 4] мәндерді таңдап, бір вектор береді.
- grid[grid[:, 0] >= 5, :] тұтас жолдарды таңдайды.

Broadcasting: пішіндердің үйлесімділігі

Пішіндер	Үйлесімді ме?	Қалай түсіндіру керек
(3, 4) және сан	Иә	Әр ұяшыққа бір сан
(3, 4) және (4,)	Иә	Әр бағанға бір мән
(3, 4) және (3, 1)	Иә	Әр жолға бір мән
(3, 4) және (3,)	Жоқ	Соңғы өлшемдер 4 және 3
(n, p) және (p,)	Иә	Мысалы, белгілердің орташа мәндері

Пішіндердің үйлесімділігі операция мағынасының дұрыстығын әлі дәлелдемейді.

Екі түрлі ось бойынша орталықтандыру

Python мысалы

```
column_means = grid.mean(axis=0)
centered = grid - column_means
row_means = grid.mean(axis=1, keepdims=True)
row_centered = grid - row_means
assert np.allclose(centered.mean(axis=0), 0)
assert np.allclose(row_centered.mean(axis=1), 0)
print(centered.shape, row_centered.shape)
```

Нәтиже: Екі нәтиженің де пішіні (3, 4), бірақ мағынасы әртүрлі.

Бағандар бойынша

- Әр белгінің орташа мәні нөл болады.
- (4,) пішініндегі вектор азайтылады.

Жолдар бойынша

- Әр жолдың орташа мәні нөл болады.
- (3, 1) пішініндегі массив азайтылады.

Reshape, транспонирлеу және сұрыптау

Пішінді өзгерту

- reshape элементтердің жалпы санын сақтайды.
- .T тек shape өзгертпей, осьтерді ауыстырады.
- flatten тәуелсіз жалпақ көшірме жасайды.

Нысандарды сақтау

- sort(axis=0) бағандарды жеке сұрыптайды.
- argsort бір белгі бойынша ретті береді.
- Бір рет жолдың барлық өрісіне қолданылады.

Пішінді ғана емес, мәндердің шығу тегін де тексеріңіз.

Дисперсия және бөлім туралы келісім

var = sum((x - mean)2) / (n - ddof)**

ddof=0

- Бөлім n: қолдағы жиынды сипаттау.
- NumPy ішіндегі әдепкі мән осы.

ddof=1

- Бөлім n-1: бағалау келісімі.
- Кітапханаларды салыстыру үшін айқын береміз.

[2, 4, 6, 8] үшін: mean=5; var(ddof=0)=5; var(ddof=1)=20/3.

std = sqrt(var); аралық мәндерді дөңгелектемейміз.

pandas кестесінен сандық матрица алу

Python мысалы

```
from pathlib import Path
import pandas as pd

football = pd.read_csv(Path('data') / 'football.csv')
features = ['Acceleration', 'SprintSpeed', 'Stamina']
matrix = football[features].to_numpy(dtype=float, copy=True)
assert matrix.shape == (12897, 3)
assert np.isfinite(matrix).all()
corr = np.corrcoef(matrix, rowvar=False)
print(corr.shape)
```

Нәтиже: Белгілер матрицасы: (12897, 3); корреляциялар: (3, 3).

Келісім

- features реті бөлек сақталады.
- copy=True кейінгі өзгерістерді оқшаулайды.

Корреляция

- rowvar=False: белгілер бағандарда.
- Сызықтық байланыс себепті анықтамайды.

Жиі кездесетін сандық қателер

Қате	Неге қауіпті	Тексеру
axis шатастыру	Басқа сұраққа жауап береді	Күтілетін пішін және мағына
view өзгерту	Бастапқы нұсқаны бұзады	Тәжірибеге дейін сору жасау
Бағандарды жеке сұрыптау	Жазбаларды бұзады	argsort және ортақ рет
* және @ шатастыру	Басқа операция	Қолмен мысал және shape
ddof ұмыту	Бөлімдер әртүрлі	Айқын параметр
Тым ерте дөңгелектеу	Қатенің жинақталуы	Тек нәтижені шығарғанда дөңгелектеу

Дәрістің негізгі нәтижелері

1. ndarray мәндермен, пішінмен және dtype арқылы беріледі.
2. axis жинақталатын осьті көрсетеді; нәтиже пішінін алдын ала болжаймыз.
3. Тілім мен көшірменің бастапқы жадпен байланысы әртүрлі.
4. Broadcasting пішіндерді сәйкестендіреді, бірақ мағынаны тексермейді.
5. Статистикалар айқын ddof пен мазмұндық түсіндіруді талап етеді.

Бақылау сұрақтары және өзіндік жұмыс

1. shape, ndim және size бір-бірінен қалай ерекшеленеді?
2. a[:, 1] және a[:, 1:2] неге әртүрлі пішін береді?
3. Таңдалған бөлікті өзгерту қашан бастапқы нұсқаны өзгертеді?
4. (n, p), (p,) және (n, 1) пішіндері қалай үйлеседі?
5. Корреляция неге себеп-салдар байланысын дәлелдемейді?

Өзіндік жұмыс

4x3 матрицасын бағандар және жолдар бойынша орталықтандырыңыз. Аралық массивтердің пішіндерін жазып, нөлдік орташа мәндерді тексеріңіз.

Дереккөздер және құжаттама

Құжаттама

1. NumPy: негіздер

https://numpy.org/doc/stable/user/absolute_beginners.html

2. NumPy: broadcasting

https://numpy.org/doc/stable/user/basics.broadcasting.html

3. NumPy: көшірмелер мен көріністер

https://numpy.org/doc/stable/user/basics.copies.html

Дәрістің негізі

- Берілген 7-дәрістің презентациясы
- Берілген 7-зертханалық жұмыс
- Берілген 7-зертханалық жұмыстың жауаптары