

تمارين سری 1 - مقدمات دوره: یادگیری ماشین

برای تمارين زیر فصل ۰ دوره، با موضوع یادگیری ماشین و کدهای مرتبط با آن را از گیت هاب زیر نیاز خواهید داشت:

<https://github.com/Alireza-Akhavan/ML-notebooks/>

سوال ۱) پیدا کردن K بهینه

در کد [طبقه بندی ارقام فارسی](#)، k بهینه در الگوریتم KNN را بیابید. برای این تمرين جست و جو در بازه ۱ تا ۱۰ کفایت میکند. لطفاً یک نمودار که محور x آن مقدار k و محور y آن دقت روی داده های X_test است را نیز رسم کرده و آن را تحلیل کنید.

سوال ۲) پیدا کردن سائز بهینه

با بهترین k که از قسمت قبل پیدا کردید، سائز بهینه تصویر به جای 5 در 5 را بیابید. توصیه میشود تمام اعداد ممکن را در ابتدا نگردید و ابتدا در بازه های بزرگتر جست و جو کنید (مثلاً s=2, 5, 10, 20, 50, 100, 200, 500) بعد بازه جست و جو را کمی کوچکتر کنید.

سعی کنید با رسم نمودارهای مقایسه آن را تحلیل کنید و بهترین size عکس را اعلام کنید.

راهنمایی: میتوانید برای انجام این کار هر بار تابع load_hoda را با پارامتر size که پیش فرض ۵ دارد فراخوانی کنید:

`load_hoda(size=5)`

سوال ۳) لود کردن تصویر و پیشبینی

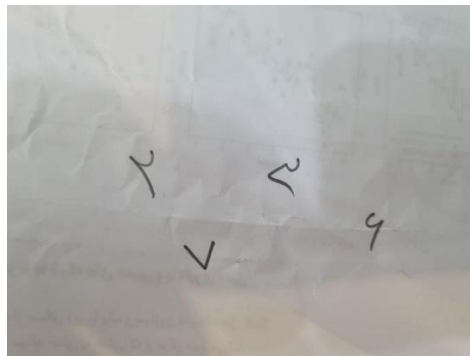
بهترین مدل خود از دو آزمایش قبلی را save کرده، سپس در یک کد دیگر بهترین مدل را لود کرده و تصویر زیر را به آن داده و پیش بینی مدل را چاپ نمایید:

<https://github.com/Alireza-Akhavan/ML-notebooks/blob/main/adad.png>

سوال ۴) تشخیص چندین عدد روی تصویر

عکس زیر که حاوی چند عدد روی صفحه هست را دانلود کرده، سپس سعی کنید از کد زیر استفاده کرده و تمامی اعداد داخل آن را پیش بینی کنید.

Code: <https://github.com/Alireza-Akhavan/SRU-deep-1402/blob/master/images/persian-digits.jpg?raw=true>



<https://github.com/Alireza-Akhavan/SRU-deep-1402/blob/master/images/persian-digits.jpg?raw=true>

سوال ۵) استفاده از مدلها خطی در تشخیص اعداد دست نویس

با فرض 5x5 بودن سائز تصاویری، آن ها را با یک مدل خطی نظیر

[LogisticRegression](#) یا [LinearSVC](#)

طبقه بندی کنید.

سپس یکبار هم

[RidgeClassifier](#)

را بررسی کنید. آیا در اینجا استفاده از این روش اوضاع را بهتر میکند؟ فکر میکنید چرا؟

در نهایت ضرایب بهترین مدل را چاپ کنید

سوال ۷) رگرسیون قیمت خانه

لطفا با یک مدل خطی رگرسیون، دیتاست قیمت خانه را با دستور زیر لود کرده و قسمتی را برای تست بردارید، در نهایت یک کدل روی آن آموزش دهید:

```
import pandas as pd
import numpy as np
```

```
file_url = "https://raw.githubusercontent.com/emanhamed/Houses-dataset/master/Houses%20Dataset/HousesInfo.txt"
columns = ["bedroom", "bathroom", "area", "zipcodes", "price"]
df = pd.read_csv(file_url, sep=" ", names=columns)
df
```

	bedroom	bathroom	area	zipcodes	price
0	4	4.0	4053	85255	869500
1	4	3.0	3343	36372	865200
2	3	4.0	3923	85266	889000
3	5	5.0	4022	85262	910000
4	3	4.0	4116	85266	971226
...

✓ مسیر مجموعه داده:

<https://raw.githubusercontent.com/emanhamed/Houses-dataset/master/Houses%20Dataset/HousesInfo.txt>