

استاد: دکتر سلیمی بدر

مهلت ۱۲/۱۰/۹۹



دانشگاه شهید بهشتی  
دانشکده مهندسی و علوم کامپیوتر

تمرین دوم کدی

مبحث یادگیری ماشین

هدف از این تمرین استفاده از الگوریتم Logistic Regression برای کلاسنندی است. دیتاست استفاده شده در این تمرین iris dataset است که تشکیل شده از اطلاعات مربوط به ۱۵۰ گل از نژادهای Iris setosa, Iris versicolor, Iris virginica میباشد.

الگوریتم Logistic Regression یک نوع دسته بند خطی میباشد که به کمک تابع سیگموئید عمل دسته بندی را انجام میدهد (توضیحات بیشتر در اسلاید های درس موجود است).

شما باید با استفاده از ویژگی های تعریف شده در دیتاست و الگوریتم مورد نظر این سه دسته را از یکدیگر جدا کنید.

دقت کنید در این تمرین استفاده از کتابخانه های آماده نظیر scikit-learn مجاز میباشد. توصیه میشود برای انجام این تمرین از زبان برنامه نویسی پایتون استفاده کنید.



با دستورات زیر میتوانید این دیتاست را به کمک scikit-learn لود کنید:

```
>> from sklearn import datasets
>> iris = datasets.load_iris()
```

فایل دیتاست نیز در اختیار شما قرار داده شده است.

همانطور که گفته شد دیتاست مورد نظر دارای سه نوع دسته میباشد. از آنجایی که الگوریتم Logistic Regression دسته بندی دوتایی انجام میدهد شما باید از یکی از سه روش زیر برای دسته بندی استفاده کنید:

One Versus One -1

One Versus Rest -2

Softmax -3 (این روش از تابع softmax که در پایین تعریف شده است برای بدست آوردن احتمال مربوط به هر دسته استفاده میکند)

$$\overline{P}_k(x) = \sigma(s(x))_k = \frac{\exp(s_k(x))}{\sum_{j=1}^k \exp(s_j(x))}$$

معیار های زیر باید برای سنجش الگوریتم مورد بررسی قرار گیرند:

Accuracy -1

Precision -2

Recall -3

توجه شود معیار های دوم و سوم برای دسته بندی دوتایی تعریف میشوند و برای استفاده در دسته بندی بیشتر از دوتا باید از یکی از روش های میانگین گیری micro, macro یا weighted استفاده شود.

(راهنمایی: برای سنجش الگوریتم میتوان از تابع cross\_val\_predict کتابخانه scikit-learn کمک گرفت. این تابع عمل تقسیم دیناست به دو قسمت train و validation را خود به خود انجام میدهد.)

سوال: کدام یک از ویژگی ها بیشترین اطلاعات را در اختیار ما قرار میدهد؟ برای پاسخ به این سوال باید دید به کمک کدام یک از ویژگی ها به تنهایی میتوان دسته بندی بهتری انجام داد.

---

توجه:

- در صورت مشاهده شباهت در کد به افراد خاطی نمره 100- تعلق می گیرد.
- در صورت داشتن هر گونه ابهام یا مشکل به آدرس [ai.99.a.ta@gmail.com](mailto:ai.99.a.ta@gmail.com) ایمیل دهید.