

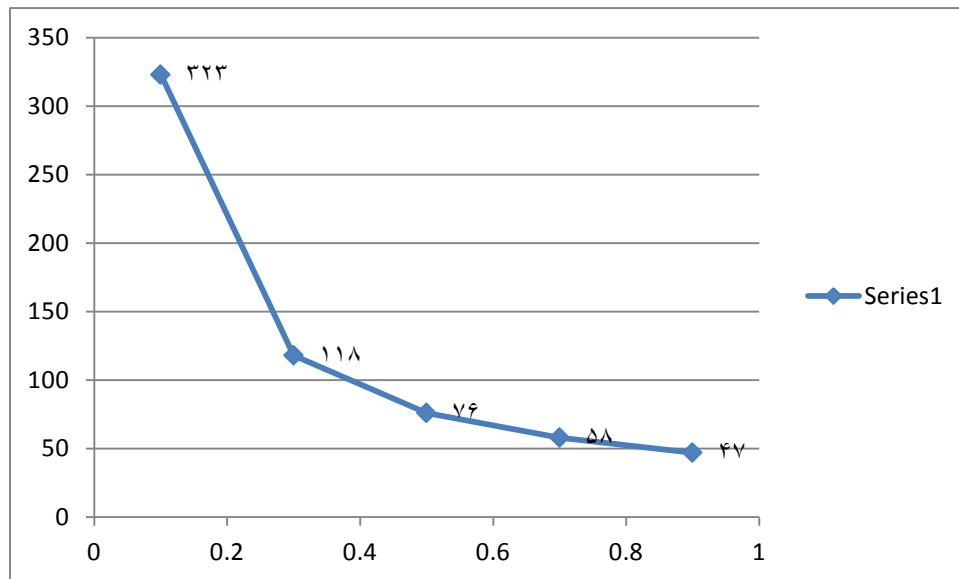
به نام خدا

تمرین سری سوم درس یادگیری ماشین

محمد میرزانژاد - ۸۳۰۵۹۶۰۲۷

۱- تغییرات تعداد تکرار با توجه به تغییر نرخهای یادگیری :

از آنجا که ارزیابی الگوریتم به صورت ۵fold است، تعداد تکرار به ازای بیشترین میزان تکرار در fold ها گزارش شده است.



به ازای هر Fold آمار به صورت ریز در برنامه ارائه شده است. در اینجا به میانگین معیارهای دقت در ۵ fold اشاره میکنیم.

نرخ یادگیری	accuracy	precision	recall	FScore	RMSE	Max Epochs
۰.۱	۹۸.۱۸	۵.۶	۱.۴	۲.۲	۰.۰۹۹	۳۲۳
۰.۳	۹۸.۱۸	۵.۶	۱.۴	۲.۲	۰.۱۰۱	۱۱۸
۰.۵	۹۸.۲۵	۵.۶	۱.۲	۱.۹۵	۰.۱۰۳	۷۶
۰.۷	۹۸.۲۵	۵.۶	۱.۲	۱.۹۵	۰.۱۰۵	۵۸
۰.۹	۹۸.۲۵	۵.۶	۱.۲	۱.۹۵	۰.۱۰۷	۴۷

توضیح برنامه :

- loadData دیتاست را میخواند
- findDatasetMinMax جهت نرمال سازی مقادیر ماکزیمم و مینیمم هر ویژگی را پیدا میکند.
- normalizeDataset مقادیر دیتاست را نرمال میکند
- crossValidationSplit دیتاست را با توجه به تعداد فولدها split یا تقسیم میکند.
- crossValidationEvaluate عمل train و test را با فراخوانی های مرتبط و با توجه به تقسیم بندی فولدها انجام میدهد. در این تابع برای هر فولد مقادیر معیارهای دقت و همچنین RMSE با فراخوان یتوابع مرتبط انجام میشود
- calculateEvaluationScores مقادیر معیارهای دقت از قبیل Precision، recall، Fscore و ... را محاسبه میکند.
- gradientDescent الگوریتم کاهش گرادیان را با توجه به نرخ یادگیری تعیین شده روی داده های آموزش اجرا میکند و شرط توقف آن خطای (دلتا) کمتر از ۰.۰۰۲ است.
- logisticRegression تابع اصلی برنامه است که با فراخوانی تابع gradientDescent مقادیر ضرایب معادله خطی را بدست می آورد و سپس تابع predict را فراخوانی میکند و نتایج پیش بینی را بر میگردداند
- predict ضرایب حاصل از آموزش را در مقدار داده تست ضرب میکند و حاصل را به تابع expit میدهد که این تابع مقدار تابع سیگموئید را برای آن محاسبه کرده و عددی بین ۰ تا ۱ خروجی میدهد ، سپس این خروجی برای تابع classify فرستاده میشود و در آنجا مقدار خروجی بسته به میزان نزدیکی به هر یک از مقادیر ۰ یا ۱ به کلاس مرتبط نسبت داده میشود.

۲- برنامه با استفاده از معادله فرم نرمال و به صورت closed form و با فرمول زیر مقدار W را محاسبه و سپس برای پیش بینی پیکسل های نیمه پایینی ، برای داده های ورودی آزمون ، ماتریس نیمه بالایی را در ماتریس W ضرب و نتیجه را به عنوان خروجی (نیمه پایینی) تصویر، به صورت فایلها ی تصویری ذخیره میکند. همچنین پیاده سازی با scikit نیز انجام شده است.

یادآور میگردد که داده های ورودی و خروجی نرمال سازی شده اند.

معیار RMSE برای داده های تست در برنامه چاپ میشود که به دلیل تعداد بالای آن (۱۰۰ عدد) در اینجا به کمترین مقدار آن برای برنامه و همچنین خروجی Scikit اشاره میکنیم :

Scikit RMSE	RMSE برنامه
۲۲	۴۵

همچنین بهترین نمونه آزمون تکمیل شده که شباهت بیشتری به نمونه اصلی و همچنین خروجی تولید شده توسط کتابخانه scikit دارد در ادامه آورده شده است.



خروجی برنامه



Scikit خروجی