



مبانی یادگیری ماشین

نیم سال دوم ۰۱-۰۲

مدرس: دکتر حامد ملک

دانشکده‌ی مهندسی کامپیوتر

موعد تحویل: ۱۴۰۲/۰۲/۰۹

تمرین سری چهارم

مسائل تئوری

مسئله‌ی ۱.

تفاوت‌های تقویت گرادیان^۱ و جنگل تصادفی^۲ را بیان کنید، عموماً کدام الگوریتم نتیجه بهتری خواهد داشت؟

مسئله‌ی ۲.

مفاهیم Boosting و Bagging و Stacking را در روش‌های Ensemble توضیح دهید.

مسئله‌ی ۳.

چرا الگوریتم تقویت گرادیان، قدرت تعمیم‌پذیری بالایی دارد؟

مسئله‌ی ۴.

در مورد AdaBoost تحقیق کرده و تفاوت آن با تقویت گرادیان را توضیح دهید.

مسئله‌ی ۵.

فرض کنید دیتاستی با ۱۰۰ ویژگی، ۱۰۰۰۰۰ نمونه و ۵ کلاس در اختیار دارید و مدل تقویت گرادیانی را با ابرپارامتر^۳ زیر آموزش داده‌اید و $f1_score = 0.5$ را بدست آورده‌اید:

```
gbc = GradientBoostingClassifier(  
    n_estimators=500,  
    learning_rate=0.05,  
    max_features=5,  
    subsample=0.3,  
    max_depth=3
```

^۱Gradient Boosting

^۲Random Forest

^۳Hyperparameters

)
هریک از ابرپارامترها و تاثیر افزایش و کاهش آن را به صورت مستقل بر روی نتیجه مدل و منحنی یادگیری تحلیل کنید.

نکات تمرین

— در صورت هرگونه تقلب نمره صفر برای شما لحاظ می گردد.

— تمرین تحویل حضوری خواهد داشت.