

## تکلیف ۶ - رگرسیون

### شرح

از کتابخانه sklearn مجموعه داده diabetes dataset را فراخوانی کنید. در این مجموعه داده اطلاعات مربوط به ۴۴۲ بیمار ثبت شده است. برای هر بیمار ۱۰ ویژگی به شرح زیر ثبت شده است. پاسخ (Target) در ستون یازدهم میزان پیشرفت دیابت بعد از یک سال به صورت عددی می باشد. هدف ایجاد مدلی برای پیش‌بینی رشد دیابت در بیماران دیابتی بر اساس ۱۰ ویژگی مزبور است. برای توضیحات بیشتر این لینک را مشاهده کنید.

[https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\\_diabetes.html#sklearn.datasets.load\\_diabetes](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_diabetes.html#sklearn.datasets.load_diabetes)

- age age in years
- sex
- bmi body mass index
- bp average blood pressure
- s1 tc, total serum cholesterol
- s2 ldl, low-density lipoproteins
- s3 hdl, high-density lipoproteins
- s4 tch, total cholesterol / HDL
- s5 ltg, possibly log of serum triglycerides level
- s6 glu, blood sugar level

برای آموزش مدل ۸۰٪ داده ها به صورت تصادفی انتخاب و اعتبار هر مدل با ۲۰٪ مابقی سنجیده شود. داده های انتخاب شده برای آموزش برای همه‌ی روش‌ها یکسان در نظر گرفته شوند. برای رگرسیون روش‌های زیر استفاده شود.

1. Linear Regression
2. Ridge Regression
3. Lasso
4. ElasticNet
5. Stochastic Gradient Descent
6. Polynomial Regression (2nd order polynomial using polynomial features)
7. Kernel ridge regression
8. Support Vector Regression
9. Nearest Neighbors Regression
10. DecisionTreeRegressor
11. AdaBoostRegressor
12. MLPRegressor

### خواسته‌ها

الف) تعیین میانگین مجذور خطای آموزش و اعتبارسنجی برای هر روش

ب) تعیین میانگین قدرمطلق خطای آموزش و اعتبارسنجی برای هر روش

ج) تعیین مقدار بهینه برای یک هایپرپارامتر در هر روش با استفاده از تابع GridsearchCV

د) مقایسه روش‌های استفاده شده