

توضیحات تکمیلی سوال ۷

(ب) هر چه معيار (R^2) به ۱ نزیک تر باشد، رگرسیون بهتری داریم. و هر چه MSE به (کمتر در واقع) صفر نزدیک تر باشد نتیجه بهتر خواهد بود. و کمتر بودن MSE داده های ارزیابی نسبت به داده های آموزشی نشانه ی **overfitting** است. در صورتی که R^2 داده های ارزیابی از R^2 داده های آموزشی کمتر باشد **overfitting** خواهیم داشت. همینطور $MSE=0$ نیز معادل **overfitting** است.

(ج) α مناسب α بی است که R^2 حداکثر و MSE حداقل باشد. (با توجه به تصادفی بودن نویز گوسی، به ازای اجرا های متفاوت نتیجه متفاوت خواهد بود. در این اجرا برای ridge ، $\alpha=1.0e-08$ ، مقدار مناسب و برای LASSO ، $\alpha=1.0e-04$ مناسب بود).

(د) هر چه مقدار α کوچکتر می شود، مقدار MSE داده های آموزشی و ارزیابی کمتر می شود. و برای هر دو رگرسیون از یک α به بعد نتیجه تغییری نمی کند (و یا بدتر می شود).

(ه) LASSO ضرایب را بیشتر **regularize** می کند. زیرا با توجه به ذات آن، یا ضرایب را نگه می دارد یا به کلی حذف می کند. ولی ridge چنین نیست و با توجه به حل مساله بهینه سازی، مقدار ضرایب را هر عددی می تواند بدست آورد. در نمودارهای مربوطه هم مشخص است که ضرایب حاصل از LASSO کوچکتر از ridge شده اند.