



## تمرین سری دوم (یادگیری بانظارت)



### الف) تمرینات classification

بنا داریم قطعات تولیدی یک کارخانه را به دو بخش سالم و خراب تقسیم کنیم. در دیتاست `quality_test.csv` قطعات سالم با برچسب ۱ و قطعات معیوب با برچسب صفر نمایش درآمده اند.

(۱) از الگوریتم KNN به منظور دسته‌بندی این داده‌ها استفاده کنید. با تغییر مقدار  $k$  بهترین دقت ممکن را به دست آورید.

(۲) از `logistic regression` به منظور دسته‌بندی داده‌های فوق استفاده کنید. در ابتدا از رگرسیون خطی استفاده کنید و سپس از رگرسیون چندجمله‌ای درجه‌ی ۲ و ۳ بهره ببرید. نتایج را با یکدیگر مقایسه کنید. چه تفاوتی را مشاهده می‌کنید؟

(۳) این بار از مدلی از درجه‌ی ۷ استفاده کنید و با استفاده از `L2-regularization` و با تعیین ضرایب ۰,۰۱ و ۱۰ و ۱۰۰۰ مجدداً جداسازی دسته‌ها را انجام دهید.

### ب) تمرینات regression

(۱) متوسط قیمت خانه در دیتاست `Boston housing` را با استفاده از رگرسیون خطی و چندجمله‌ای از درجه‌ی ۲ و ۳ پیش‌بینی کنید. خطای `MSE` مدل‌ها را به دست آورده و در جدولی با یکدیگر مقایسه نمایید. (حتماً داده‌ها نرمالایز شوند).

(۲) در طول کلاس الگوریتم رگرسیون خطی بر روی دیتاست بارهای حرارتی ساختمان (`ENB2012_data.csv`) به کار برده شد. در این تمرین قصد بر آن است که رگرسیون غیرخطی بر روی این داده‌ها انجام شده و نتایج آن‌ها با حالت قبل مقایسه شود.

ابتدا داده‌ها را نرمالایز کنید و سپس مدلی از درجات ۲ و ۳ بر روی آن‌ها برازش کنید. در پایان `R2-score` و `MSE` را با رگرسیون خطی مطرح شده در کلاس مقایسه کنید. آیا نتایج در مقایسه با حالت رگرسیون خطی بهتر هستند یا بدتر؟

لطفا هر یک از بخش‌ها را در یک فایل jupyter notebook انجام داده و ارسال نمایید. جدول‌های مقایسه هم در یک فایل pdf ارسال شوند. در گزارش درج میزان دقت هر یک از مدل‌ها داخل جدول کفایت می‌کند و نیازی به توضیح راجع به آن‌ها نیست.