

## تمارین یادگیری ماشین سری شماره یک

### تمرین ۱

از داده های ضمیمه شده به دلخواه سه مورد را انتخاب کنید. داده ها را به ۸۰ درصد داده آموزشی و ۲۰ درصد داده آزمایشی تقسیم کنید و روش های زیر را اعمال کنید.

الف) از روش های کلاس بندی مبتنی بر رگرسیون خطی با تابع های هزینه MAE و MSE برای کلاس بندی این داده ها استفاده کنید. دقت آزمایش (Test) و آموزش (Train) را گزارش کنید.

ب) همچنین از روش های Ridge و Lasso نیز برای اینکار استفاده نمایید. ← از چه روشی برای تخمین بهترین پارامتر می توان استفاده کرد؟ دقت این روشها را برحسب معیار (Accuracy) گزارش کنید.

ج) از روش Logistic Regression Classifier برای کلاس بندی این داده ها استفاده کنید. نتایج را با روش های قبلی مقایسه کنید.

### تمرین ۲

داده های فوق را در نظر بگیرید.

$$X = [X_1, X_2, \dots, X_N] \in \mathbb{R}^{D \times N}$$

که D تعداد ویژگی ها و N تعداد نمونه های مربوط به داده آموزشی باشد، آنگاه روش فوق را برای کلاس بندی داده ها به کار ببرید. اگر Y داده آزمایشی (تست) باشد در این صورت:

$$\min \|Y - XW\|_2^2, \|W\| < k \oplus$$

که k برابر ۳۰٪ داده های آموزشی است.

مسئله  $\oplus$  توسط OMP می تواند حل شود. بعد از حل این مسئله عنصر i ام W یعنی  $W_i$  میزان تاثیر داده  $X_i$  در ساختن Y را نشان می دهد. چگونه می توان از این  $W_i$  ها برای کلاس بندی استفاده کرد؟ پیاده سازی (کدنویسی) کنید.

این مسئله را برای داده های ORL نیز انجام دهید.