



## پیاده سازی الگوریتم‌های پایه ماشین لرنینگ

هدف اصلی این تمرین پیاده سازی دو الگوریتم پایه از دو دسته الگوریتم‌های نظارتی و غیرنظارتی به کمک کتابخانه یادگیری ماشین در پایتون (Sickit-Learn) است.

ابتدا فایل اولیه BasicMachineLearningAlgorithms.ipynb برای شروع این تمرین را دانلود کنید. وظیفه شما برای این تمرین، کامل کردن کدهای داخل این فایل است. به این صورت که ۱۰ قسمت اصلی در این فایل قرار دارد و شما می‌بایستی فقط قسمت‌هایی که با *Your code here* مستندسازی شده‌اند را کامل کنید.

لطفا به نکات زیر حتما توجه کنید:

۱. در باقی کدهایی که در این فایل نوشته شده‌اند و امضای توابع نوشته شده در این فایل تغییری ایجاد نکنید.
۳. پس از تکمیل کدها، تمامی توابع پیاده‌سازی شده (آنهایی که با *Your code here* مستندسازی شده‌اند) را داخل یک فایل پایتون به نام `Utils.py` کپی پیست کنید. دقت فرمایید که در صورت انجام ندادن این کار نمره‌ای دریافت نخواهید کرد.
۴. تمیز بودن و زیبا و خوانا بودن نمودارهایی که رسم می‌کنید به شدت تاثیر مستقیمی در نمره شما خواهد داشت.
۵. فایل نهایی که آپلود می‌کنید باید تنها شامل دو فایل باشد: `BasicMachineLearningAlgorithms.ipynb` و `Utils.py`.

## بخش اول – رگرسیون

سوالات و پیاده سازی‌هایی که باید در این قسمت انجام دهید به ترتیب به شرح زیر است:

۱. پیاده سازی یک رگرسیون خطی روی دیتاست داده شده. (۲۰ نمره)
۲. تحلیل مدل آموزش دیده شده و بررسی پارامترهای بدست آمده از مدل (۵ نمره)
۳. رسم تابع بدست آمده به همراه نقاط دیتا (۵ نمره)
۴. پیاده سازی توابع هزینه مختلف مناسب برای مساله رگرسیون و بررسی فواید آنها نسبت به یکدیگر (۵ نمره)
۵. پیاده سازی رگرسیون چندجمله‌ای (polynomial regression) و بررسی تعمیم پذیری مدل‌ها متفاوت با درجه‌های چندجمله‌ای متفاوت (موضوع underfitting vs. overfitting) (۱۵ نمره)

6. پیدا کردن بهترین مدل با بررسی روی داده تست و ارزیابی آنها (۱۰ نمره)
7. رسم مدل های نهایی بدست آمده (۵ نمره)

## بخش دوم – الگوریتم خوشه بندی K-means

سوالات و پیاده سازی هایی که باید در این قسمت انجام دهید به ترتیب به شرح زیر است:

1. پیاده سازی الگوریتم K-means روی دیتاست داده شده (۱۵ نمره)
2. رسم خوشه های بدست آمده به همراه مرکز (centroids) های آنها (۱۵ نمره)
3. تحلیل نمودار Elbow برای بدست آوردن بهترین مقدار برای تعداد خوشه های این مساله (۱۰ نمره).

توضیحات تکمیلی در فایل نوت بوک به صورت واضح و کامل آمده است.

## دیتاست

در این تمرین از دو دیتاست استفاده شده است که یکی برای بخش اول و دیگری برای بخش دوم است.

- دیتاست insurance.csv: این دیتاست شامل اطلاعات کلی از بیماران یک بیمارستان است که ستون های این دیتا به شرح زیر است:

age	sex	bmi	smoker	children	charges
سن شخص	جنسیت	شاخص توده بدنی	سیگاری است یا خیر	تعداد فرزندان	حق بیمه

هدف از این مساله پیش بینی حق بیمه براساس یکی از ویژگی های ذکر شده است.

- دیتاست mall\_customers.csv: این دیتاست شامل اطلاعات کلی از مشتریان یک مرکز خرید است که ستون های این دیتا به شرح زیر است:

age	sex	annual_income	spending_score
سن شخص	جنسیت	درآمد سالانه	توانایی مشتری برای خرید

هدف از این مساله خوشه بندی مشتریان این مرکز خرید براساس یکی از ویژگی های ذکر شده است برای استفاده های آینده اعم از ارائه سرویس های مشتری مداری بهتر و customize شده است.

لطفاً به نکات زیر توجه فرمایید:

- فرمت نام‌گذاری فایلی که آپلود می‌کنید حتماً به صورت **[student name][student id]** باشد.
- در صورت مشاهده هرگونه تقلب نمره **صفر** برای تکلیف در نظر گرفته می‌شود.
- می‌توانید سوالات و ابهامات خود را از **mohammad99hashemi@gmail.com** بپرسید.