



آزمایشگاه هوش مصنوعی کاربردی  
دانشکده مهندسی مکانیک  
دانشگاه تهران



# دوره آشنایی با هوش مصنوعی

## پروژه بخش یادگیری ماشین

استاد مربوطه:

دکتر شریعت پناهی

مسئول آزمایشگاه:

دکتر کاشانی

دستیاران آموزشی:

فاطمه مجاب

سعید دلیر

پاییز ۱۴۰۱

## بخش اول: پیش‌بینی بارهای گرمایشی و سرمایشی یک ساختمان

در این بخش می‌خواهیم بارهای گرمایشی و سرمایشی یک ساختمان را بر پایه ویژگی‌های هشت‌گانه آن ساختمان پیش‌بینی کنیم. برای این کار ابتدا مجموعه داده مورد نظر را از این [لینک](#) دانلود کنید و با ویژگی‌های داده‌ها آشنا شوید. سپس با استفاده از ویژگی‌های ۸ ستون اول ( $X_1$  تا  $X_8$ ) مقدار بارهای گرمایشی و سرمایشی ساختمان ( $Y_1$  و  $Y_2$ ) را با استفاده از رگرسیون چند متغیره پیش‌بینی کنید.

در بسیاری از مسائل رگرسیون به علت پیچیده بودن رابطه بین ورودی و خروجی‌ها به جای رگرسیون خطی از رگرسیون غیرخطی (چند جمله‌ای) استفاده می‌شود. این موضوع در مسائل دسته‌بندی نیز با غیرخطی شدن مرز تصمیم‌گیری خود را نشان می‌دهد.

در این گونه مسائل به جای استفاده از رابطه خطی، از یک چند جمله‌ای استفاده می‌کنیم و در واقع فضای ویژگی‌ها را به فضای مرتبه بالاتری تغییر می‌دهیم. اگر فضای ویژگی‌ها را به درجه  $d$  ببریم، تمام جملات چند جمله‌ای "حداکثر" از درجه  $d$  در فضای ویژگی‌ها وجود خواهند داشت. به عنوان مثال اگر برای رگرسیون دو متغیره فضای ویژگی‌ها را به درجه ۳ ببریم فضای ویژگی‌ها به صورت زیر خواهد بود.

$$1, x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, x_1x_2^2, x_2^3$$

### مراحل اجرایی و خواسته‌ها:

- (۱) ضرورت Normalize کردن داده‌ها را به اختصار توضیح دهید و ستون‌های  $X_1$  تا  $X_8$  (feature ها) را Normalize کنید.
- (۲) یک مدل خطی روی مجموعه داده‌ها نرمالایز شده برازش کرده و مقادیر mean squared error و  $r^2$  score را گزارش کنید.
- (۳) یک مدل غیرخطی از درجه ۳ روی مجموعه داده‌ها نرمالایز شده برازش کرده و مقادیر mean squared error و  $r^2$  score را گزارش کنید. اگر از مجموعه داده‌ها اصلی (بدون نرمالایز کردن) استفاده کرده بودیم این مقادیر به چه صورت تغییر می‌کردند؟

## توضیحات

- برای حل و پیاده‌سازی سوالات تنها استفاده از زبان برنامه‌نویسی پایتون مجاز است. همچنین شما مجاز به استفاده از کتابخانه‌های آماده مانند numpy, matplotlib, pandas و sklearn می‌باشید.
- تحویل گزارش برای این پروژه ضروری است. لزومی به توضیح جزئیات کدها نیست اما از آنجا که برای این پروژه از کتابخانه‌های آماده استفاده می‌شود، لطفاً تمامی پارامترهای تنظیم‌شده در هر قسمت از کد را گزارش کرده و فرض‌هایی را که برای پیاده‌سازی‌ها و محاسبات خود به کار برده‌اید ذکر کنید. ضمناً در فرایند ارزیابی کدهای شما لزوماً اجرا نخواهند شد، بنابراین همه‌ی نتایج و تحلیل‌های خود را به‌طور کامل ارائه کنید.
- گزارش شما باید به صورت تایپ شده و با فرمت pdf ارائه شود و کدهایی که به همراه گزارش تحویل می‌دهید باید قابل اجرا باشند. در انتها تمامی فایل‌های لازم را در یک فایل zip یا rar بارگذاری و برای دستیاران آموزشی ارسال کنید.
- پس از انجام و تحویل این تمرین، پاسخ خود را به دستیار آموزشی ارسال کرده و تمرین بخش بعدی را دریافت کنید.
- پرسش‌های خود را از طریق تلگرام از دستیاران آموزشی مربوطه بپرسید:

@saeed675

@SFatemehM