

به نام خالق عقل و خرد



## جلسه دوم

یادگیری ماشین

نیم سال دوم ۱۴۰۴ - ۱۴۰۳

دانشکده ریاضی و آمار

دانشگاه اصفهان

---

اسفند ماه ۱۴۰۳

استاد درس:

دکتر فاطمه منصوری

## عنوان تمرین : پیاده‌سازی رگرسیون چند جمله‌ای و روش‌های Regularization

هدف: درک و پیاده‌سازی polynomial regression و اجرای تکنیک‌ها مختلف Regularization و بررسی نتایج

دستورالعمل :

۱. یک دیتاست ساختگی ایجاد کنید. برای این کار یک تابع بنویسید که (عنوان تابع در فایل ژوپیتتر قرار داده شده است) دیتاست را مطابق زیر ایجاد کند. دیتاست شامل ۱۰۰۰ نمونه باشد. دیتاست شامل یک ویژگی مستقل  $x$  است. مقدار  $x$  بین ۳- تا ۳ تغییر می‌کند و ارتباط بین متغیر وابسته و متغیر ساختگی مطابق زیر باشد.

$$y = 0.2x^3 - 0.5x^2 + 3x + 10 + noise$$

۲. با استفاده از یک scatter plot ارتباط بین  $x$  و  $y$  را مشاهده نمایید.

۳. داده‌ها را به train و test تقسیم کنید. علاوه بر آن یک مجموعه تست در بازه ۵ تا ۷ به نام out\_of\_range\_test ایجاد کنید.

۴. برای ویژگی polynomial feature ایجاد کنید و یک مدل polynomial regression به داده‌ها اختصاص دهید.

۵. تابع train\_models را مطابق آنچه در فایل ژوپیتتر خواسته شده است پیاده‌سازی کنید و سپس با استفاده از آن مدل‌ها با درجات مختلف ایجاد کرده و نمایش دهید.

۶. تابع evaluate\_models نوشته شده در فایل ژوپیتتر را کامل کنید به صورتیکه به ازای مدل‌های ساخته شده با درجات مختلف مقدار خطا را محاسبه کرده و در آخر لیستی از خطاها برگرداند.

۷. روش‌های مختلف regularization (lasso, ridge) را برای مدل با درجه ۶ امتحان کنید.

۸. ضرایب به دست آمده در هر روش regularization را با استفاده از نمودار میله‌ای نمایش داده و از نظر sparsity با یکدیگر مقایسه کنید.

۹. در مراحل قبل بهترین درجه را برای دیتاست ساختگی که ریسک overfitting ندارد به دست آوردید. در این مرحله بهترین مقدار alpha را برای مدل با آن درجه به دست آورید.