



به نام خدا  
هوش محاسباتی (بهار ۱۴۰۱)

تکلیف شماره ۲ (کامپیوتری) : Logistic Regression/Neural Networks

موعد تحویل: ۲ اردیبهشت ماه

در این تکلیف قصد داریم عملکرد روش‌های رگرسیون لاجیستیک و شبکه‌های عصبی چند لایه را در مساله دسته‌بندی با هم مقایسه کنیم. بدین منظور مساله تشخیص سطح اوزون در جو، را در نظر می‌گیریم. برای این کار از یک مجموعه داده شامل ۲۵۳۶ داده و ۷۳ ویژگی که هر یک ساعت جمع‌آوری شده، استفاده می‌کنیم. (برای دریافت اطلاعات بیشتر در مورد این مجموعه داده می‌توانید به [این لینک](#) مراجعه فرمایید)

- مجموعه داده در فایل Ozone Level Detection.data موجود می‌باشد.
- انجام این تمرین به صورت فردی می‌باشد.
- حتما پیاده‌سازی‌های خود را در محیط Jupyter Notebook و در یک فایل ipynb انجام دهید.
- برای پیاده‌سازی بخش شبکه عصبی، از یکی از کتابخانه‌های PyTorch، TensorFlow یا Keras استفاده نمایید.
- برای پیاده‌سازی بخش رگرسیون لاجیستیک مجاز به استفاده از کتابخانه‌های ذکر شده نیستید.
- نیازی به یک فایل پی دی اف جداگانه برای گزارش کار نیست. توضیحات خود را در همان فایل ipynb بنویسید. توضیحات به فارسی نوشته شوند.

۳	پیش پردازش (رفع Missing value) و پیاده سازی kfold cross validation
۳	پیاده سازی مدل رگرسیون لاجیستیک
۳	ارزیابی مدل رگرسیون لاجیستیک (محاسبه دقت، گزارش میانگین و انحراف معیار به ازای foldهای مختلف)
۱/۵	بررسی وجود پدیده overfitting و underfitting در مدل رگرسیون لاجیستیک
۳	پیاده سازی مدل شبکه عصبی
۳	ارزیابی مدل شبکه عصبی (محاسبه دقت، گزارش میانگین و انحراف معیار به ازای foldهای مختلف)
۲	گزارش نتایج شبکه عصبی برای مقادیر گوناگون هایپرپارامترها
۱/۵	بررسی وجود پدیده overfitting و underfitting در مدل شبکه عصبی

### پیش پردازش:

همانطور که در مجموعه داده مشاهده می کنید، تعدادی داده از دست رفته (Missing values) وجود دارد. در مورد روش های حل این مشکل جستجو کرده و بهترین روش برای مجموعه داده خود را انتخاب و پیاده سازی نمایید.

همچنین برای تقسیم بندی داده ها به دو بخش آموزش و تست، روش kfold cross validation با مقدار  $K=3$  را به کار برده و برای دقت بدست آمده از حالت های مختلف آن، مقدار میانگین و انحراف معیار را گزارش کنید.

### رگرسیون لاجیستیک:

در این قسمت از تمرین قصد داریم با استفاده از روش رگرسیون لاجیستیک سطح اوزون را پیش بینی کنیم. بعد از انجام بهینه سازی تابع هزینه و انتخاب بهترین پارامترها با روش Gradient Descent، سعی کنید با انتخاب

مقادیر مختلف پارامتر regularization، دو پدیده overfitting و underfitting را در یادگیری پارامترها توضیح دهید.

برای ارزیابی مدل خود از معیار Accuracy استفاده نمایید.

### شبکه‌های عصبی:

در این مرحله سعی داریم که مساله پیش‌بینی مورد نظر را با استفاده از شبکه عصبی چند لایه انجام دهیم. شبکه شما ممکن است ساختارهای گوناگونی از نظر تعداد لایه‌ها، تعداد نرون‌های هر لایه، نرخ یادگیری و ... داشته باشد. بنابراین حتما مقادیر مختلفی برای این پارامترها تست کرده و نتایج حاصل را با یکدیگر مقایسه و گزارش کنید. برای ارزیابی مدل خود از معیار Accuracy استفاده نمایید.

نشان دهید که مدل‌تان دچار مشکلات Underfit یا Overfit نشده است.

**موفق باشید**