

# دوره یادگیری عمیق

تمرین ۵

## پروژه MLP

در این تکلیف، شما با دو مجموعه داده مختلف کار خواهید کرد. وظیفه شما پیاده‌سازی شبکه‌های عصبی چندلایه (MLP) با استفاده از PyTorch برای حل یک مسئله رگرسیون و یک مسئله طبقه‌بندی است.

### مجموعه داده ۱: عملکرد مطالعاتی دانش آموزان

مجموعه داده اول، مجموعه داده "عملکرد مطالعاتی دانش آموزان" است. این مجموعه داده شامل اطلاعاتی درباره عادات مطالعاتی دانش آموزان و عملکرد آن‌ها در مدرسه است. شما می‌توانید مجموعه داده را از این [پیوند](#) دانلود کنید. وظیفه شما پیش‌بینی عملکرد دانش آموزان بر اساس عادات مطالعاتی آن‌ها است. این یک مسئله رگرسیون است.

### ویژگی‌ها

در این مجموعه داده پنج ویژگی مختلف وجود دارد:

- جنسیت: جنسیت دانش آموزان - (مرد/زن)
  - نژاد/اقوام: نژاد/اقوام دانش آموزان - (گروه A، B، C، D، E)
  - سطح تحصیلات والدین: آخرین سطح تحصیلی والدین - (دیپلم، لیسانس، کارشناسی، کارشناسی ارشد، دکترا)
  - ناهار: داشتن ناهار قبل از آزمون (معمولی یا رایگان/کاهش یافته)
  - دوره آمادگی آزمون: کامل یا ناقص بودن قبل از آزمون
- سه متغیر هدف وجود دارد که مدل شما باید پیش‌بینی کند:
- نمره ریاضی
  - نمره خواندن
  - نمره نوشتن

### وظایف

- پیش‌پردازش داده: مجموعه داده را بارگیری کرده و هر گونه مراحل پیش‌پردازش لازم را انجام دهید. این ممکن است شامل رفع مشکل مقادیر گم‌شده، کدگذاری متغیرهای دسته‌ای، و غیره باشد.
- ساخت مدل: یک MLP با استفاده از PyTorch پیاده‌سازی کنید. شبکه شما باید حداقل یک لایه پنهان داشته باشد. می‌توانید تعداد نوروں‌ها را در لایه(لایه‌های) پنهان انتخاب کنید.
- آموزش: مدل خود را با استفاده از یک تابع از دست دادن مناسب برای رگرسیون آموزش دهید. تابع از دست دادن را برای اطمینان از یادگیری مدل‌تان نظارت کنید.
- ارزیابی: عملکرد مدل خود را بر روی یک مجموعه داده آزمایشی جداگانه ارزیابی کنید.

## مجموعه داده ۲: سطوح چاقی

مجموعه داده دوم، مجموعه داده "سطوح چاقی" است. این مجموعه داده شامل اطلاعاتی درباره عادات غذایی و وضعیت فیزیکی افراد است. شما می‌توانید مجموعه داده را از این [پیوند](#) دانلود کنید. وظیفه شما پیش‌بینی سطح چاقی بر اساس ویژگی‌های داده شده است. این یک مسئله طبقه‌بندی است.

### ویژگی‌ها

این مجموعه داده شامل 16 ویژگی است، و این‌ها ویژگی‌ها عبارتند از:

- **جنسیت:** جنسیت فرد.
  - **سن:** سن فرد.
  - **قد:** قد فرد به متر.
  - **وزن:** وزن فرد به کیلوگرم.
  - **سابقه خانوادگی با چاقی:** آیا فرد سابقه چاقی در خانواده دارد یا نه.
  - **FAVC (مصرف مکرر غذاهای با انرژی بالا):** آیا فرد مکرراً غذاهای با انرژی بالا مصرف می‌کند یا نه.
  - **FCVC (فراوانی مصرف سبزیجات):** چقدر فرد به طور مداوم سبزیجات مصرف می‌کند.
  - **NCP (تعداد وعده‌های اصلی):** تعداد وعده‌های غذایی که فرد روزانه مصرف می‌کند.
  - **CAEC (مصرف غذا بین وعده‌ها):** چقدر فرد بین وعده‌های غذایی می‌خورد.
  - **سیگار:** آیا فرد سیگار می‌کشد یا نه.
  - **CH2O (مصرف روزانه آب):** مقدار آبی که فرد روزانه مصرف می‌کند.
  - **SCC (نظارت بر مصرف کالری):** آیا فرد مصرف کالری خود را نظارت می‌کند یا نه.
  - **FAF (فراوانی فعالیت‌های فیزیکی):** فراوانی فعالیت‌های فیزیکی فرد.
  - **TUE (زمان استفاده از دستگاه‌های فناوری):** زمانی که فرد صرف استفاده از دستگاه‌های فناوری می‌کند.
  - **CALC (مصرف الکل):** فراوانی مصرف الکل توسط فرد.
  - **MTRANS (حمل و نقل استفاده شده):** حالت حمل و نقلی که فرد استفاده می‌کند.
- متغیر هدف **NOBeyesdad** است که سطح چاقی فرد را نشان می‌دهد. این هفت مقدار را اختیار می‌گیرد: وزن کافی، وزن نرمال، میزان چاقی مرحله I، میزان چاقی مرحله II، چاقی نوع I، چاقی نوع II و چاقی نوع III.

### وظایف

1. **پیش‌پردازش داده:** مجموعه داده را بارگیری کرده و هر گونه مراحل پیش‌پردازش لازم را انجام دهید. این ممکن است شامل رفع مشکل مقادیر گم‌شده، کدگذاری متغیرهای دسته‌ای، و غیره باشد.
2. **ساخت مدل:** یک MLP با استفاده از PyTorch پیاده‌سازی کنید. شبکه شما باید حداقل یک لایه پنهان داشته باشد. می‌توانید تعداد نوروں‌ها را در لایه(لایه‌های) پنهان انتخاب کنید.
3. **آموزش:** مدل خود را با استفاده از یک تابع از دست دادن مناسب برای طبقه‌بندی آموزش دهید. تابع از دست دادن را برای اطمینان از یادگیری مدل‌تان نظارت کنید.
4. **ارزیابی:** عملکرد مدل خود را بر روی یک مجموعه داده آزمایشی جداگانه ارزیابی کنید. می‌توانید از دقت، دقت، بازخوانی، امتیاز F1، و غیره به عنوان معیارهای ارزیابی خود استفاده کنید.