



دانشگاه صنعتی اصفهان  
دانشکده مهندسی برق و کامپیوتر

درس مبانی هوش محاسباتی

تکلیف کامپیوتری اول

تاریخ تحویل: ۱۷ فروردین

## سؤال ۱

بدون استفاده از کتابخانه های Pytorch و Tensorflow تابعی به نام `logistic_regression()` بنویسید که ورودی های زیر را بگیرد:

- $x$  : یک آرایه با اندازه  $(n\_samples, n\_features)$  که به ترتیب تعداد داده های ورودی و تعداد ویژگی های آن هستند.

- $y$  : یک وکتور به اندازه  $n\_samples$  که شامل برچسب های داده های ورودی ( $x$ ) است.

- `learning_rate` : که نرخ یادگیری است.

- `num_iterations` : که تعداد تکرارهای الگوریتم است.

- `add_intercept` : که یک متغیر بولین برای تعیین دجود یا عدم وجود `bias term` است.

داخل این تابع باید رگرسیون لاجیستیک را با نزول گرادیان پیاده سازی کنید تا وزن های یک مسئله ی دسته بندی را یادبگیرد. خروجی شما باید یک وکتور با اندازه  $n\_features+1$  باشد که شامل وزن های خروجی و یک `intercept_term` در صورتی که `add_intercept=True` باشد است.

تابع خود را با توجه به نکات زیر و با استفاده از دیتاست موجود در این [صفحه](#) آموزش دهید و تست کنید. از داده های `clean2` برای آموزش و از داده های `clean1` برای تست استفاده کنید. این دیتاست شامل ۱۶۶ نوع اطلاعات در مورد بیش از شش هزار نوع مولکول مختلف هستند که در دو دسته ی بودار و بی بو قرار گرفته اند.

- داده های خود را پیش پردازش کنید (نرمالیز کردن داده ها).

- مدل خود را آموزش و تست کنید. سپس این کار را با نرخ یادگیری کوچک و بار دیگر با نرخ یادگیری بزرگ تکرار کنید. و نتایج را مقایسه کنید.

- در هر حالت با استفاده از رسم نمودار عملکرد مدل خود را تصور کنید.

- در هر حالت عملکرد مدل خود را بدون استفاده از کتابخانه و با محاسبه ی پارامترهای دقت و خطا ارزیابی کنید.

- (امتیازی) مدل خود را با استفاده از پارامترهای `Accuracy`, `Recall` و `F1 score` ارزیابی کنید (میتوانید از توابع کتابخانه ها استفاده کنید). هر کدام از این پارامترهای نمایانگر چیست؟ توضیح دهید. (برای این کار علاوه بر فایل نوت بوک، فایلی حاوی توضیحات خود در سامانه آپلود کنید. دقت کنید که در صورت ارائه ندادن توضیحات برای این قسمت نمره ی آن برای شما در نظر گرفته نخواهد شد).

بارم بندی سوال ۱	
۵	پیش پردازش داده های ورودی
۱۰	پیاده سازی تابع رگرسیون لاجیستیک
۱۰	آموزش و تست مدل
۱۰	بررسی مدل با استفاده از نمودار
۵	محاسبه ی دقت و خطای مدل
۱۰ (امتیازی)	محاسبه ی پارامترهای ذکر شده و توضیح آنها

## سؤال ۲

در این تمرین قصد داریم یک مدل شبکه عصبی سه لایه (شامل لایه ورودی، لایه‌های مخفی و لایه خروجی) را پیاده سازی کنیم برای این کار شما مجاز به استفاده از توابع کتابخانه‌های آماده مثل pytorch و tensorflow نیستید. می‌خواهیم به کمک مدل طراحی شده یک دسته‌بندی<sup>۱</sup> بر روی مجموعه داده iris انجام دهیم. این مجموعه داده دارای چهار ویژگی (طول و عرض کاسبرگ و گلبرگ) برای ۱۵۰ نمونه است که شامل ۳ دسته مختلف هستند. در جدول زیر چند رکورد از این مجموعه داده قابل مشاهده است.

sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5.0	3.6	1.4	0.2

داده‌ها را می‌توانید از کتابخانه [scikit-learn](#) بخوانید.

(الف) ابتدا داده‌ها را به صورت تصادفی دو دسته آموزش<sup>۲</sup> و آزمایش<sup>۳</sup> تقسیم کنید. (۸۰ درصد داده‌ها را برای آموزش و ۲۰ درصد را برای آزمایش استفاده کنید.) برای تقسیم‌بندی داده‌ها مجاز به استفاده از توابع آماده نیستید. (۲ نمره)

(ب) تابعی با نام train\_SGD تعریف کرده، این تابع باید مجموعه داده، نرخ یادگیری<sup>۴</sup> و تعداد اپاک را به عنوان ورودی دریافت کرده، و فرآیند آموزش شبکه را به روش SGD انجام دهد. برای ساده کردن کار خود می‌توانید برای هر یک از قسمت‌های انتشار رو به جلو<sup>۵</sup>، پس انتشار خطا<sup>۶</sup> محاسبه خطا و ... تابع جداگانه نوشته و آن توابع را در train\_SGD فراخوانی کنید. پس از آموزش مدل بر روی داده (۸ نمره)

(ج) تابعی با نام train\_minibatchGD تعریف کرده، این تابع باید مجموعه داده، نرخ یادگیری<sup>۷</sup> و تعداد اپاک را به عنوان ورودی دریافت کرده، و فرآیند آموزش شبکه را به روش mini batch GD انجام دهد. (۸ نمره)

(د) تابعی با نام train\_GD تعریف کرده، این تابع باید مجموعه داده، نرخ یادگیری<sup>۸</sup> و تعداد اپاک را به عنوان ورودی دریافت کرده، و فرآیند آموزش شبکه را به روش GD انجام دهد. (۸ نمره)

(ه) مدلی که در قسمت د بر روی داده آموزشی، آموزش دید را بر روی داده آزمایش بررسی کنید و نمودارهای خطا بر حسب اپاک و دقت بر حسب اپاک را رسم کنید. (۸ نمره)

<sup>1</sup>Classification

<sup>2</sup>Train data

<sup>3</sup>Test data

<sup>4</sup>Learning rate

<sup>5</sup>Feed forward

<sup>6</sup>Back propagation

<sup>7</sup>Learning rate

<sup>8</sup>Learning rate

و) صحت فرمول‌های مشتق‌گیری را با روشهای عددی نشان دهید. (۸ نمره)

ز) مدل قسمت "د" را با سه نقطه شروع متفاوت بر روی داده آموزشی آموزش داده و نتایج سه حالت را مقایسه کنید. (۸ نمره)

## نکات تکمیلی

۱. برای انجام این تکلیف استفاده از زبان پایتون الزامی است.
۲. تکالیف را در محیط jupyter notebook پیاده‌سازی کنید و فایل ipynb را ارسال کنید.
۳. توضیح کدی که نوشته‌اید، بررسی و تحلیل نتایج آن و بیان علت نتایج و نیز مقایسه نتیجه با آنچه مورد انتظارتان بوده است، از اهمیت بالایی برخوردار است. شما می‌توانید گزارش پروژه را در همان محیط jupyter notebook بنویسید و نیازی به فایل pdf جداگانه نیست. هم‌چنین اگر برای حل سوال فرضیات خاصی مدنظر دارید حتماً آن را در متن گزارش قید کنید.
۴. فرمت نامگذاری تکلیف ارسالی باید به صورت زیر باشد: HWX\_Programming\_LastName\_StudentID که X شماره تکلیف LastName نام خانوادگی شما و StudentID شماره دانشجویی شما است.
۵. انجام این تکلیف به صورت تک نفره است. در صورت مشاهده تقلب، نمرات هم مبدا کپی و هم مقصد آن صفر لحاظ می‌شود.
۶. شما می‌توانید تا یک هفته پس از پایان مهلت تکلیف آن را در یکتا بارگذاری کنید. در این صورت به ازای هر روز تاخیر ۵ درصد از نمره تکلیف کسر می‌شود. پس از اتمام این یک هفته امکان ارسال با تاخیر وجود ندارد.
۷. در صورت وجود هر گونه ابهام و یا سوال می‌توانید سوالات خود را در گروه تلگرام بپرسید. هم‌چنین می‌توانید برای رفع ابهامات با دستیاران آموزشی از طریق تلگرام در تماس باشید.

آیدی‌ها:

Fatemeh\_1241

alirez anum1

alireza20010226

موفق باشید.