

به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



شبکه‌های عصبی مصنوعی و یادگیری عمیق

تمرین شماره ۲

فروردین ۱۴۰۰

فهرست سوالات

- سوال ۱ – MLP (Regression) ۳
- سوال ۲ MLP (Classification) ۴
- سوال ۳ Dimension Reduction ۶

سوال ۱ – MLP (Regression)

هدف این تمرین تعیین قیمت خانه برحسب ویژگی‌های آن است. اطلاعات مربوط به این داده‌ها در فایل data_description.txt و خود داده‌ها در فایل house_prices.csv قرار دارند.

الف) ابتدا پیش‌پردازش‌های لازم برای آماده‌سازی دیتا را انجام دهید و آن‌ها را توضیح دهید.

ب) به‌صورت تصادفی ۸۰ درصد داده‌ها را به عنوان داده آموزشی و ۲۰ درصد را به عنوان داده تست در نظر بگیرید. شبکه عصبی چند لایه‌ای را طراحی کنید که قیمت خانه را پیش‌بینی کند. تعداد لایه و تعداد نوروں و تابع فعال‌ساز مناسب برای هر لایه را مشخص کنید و آن‌ها را تحلیل کنید. (دو حالت برای تعداد لایه و دو حالت برای تابع فعال‌ساز در نظر بگیرید).

ج) با ثابت در نظر گرفتن پارامترهای به دست آمده در قسمت (ب) و با در نظر گرفتن MSE به عنوان تابع loss، مقادیر metric های MSE و MAE را در هر ایپاک برای داده آموزشی و داده تست در یک نمودار رسم کنید (یک نمودار برای معیار MSE و یک نمودار برای MAE). برای داده تست نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی را رسم نمایید. تعداد ایپاک بهینه در این قسمت را مشخص کنید.

د) با ثابت در نظر گرفتن پارامترهای به دست آمده در قسمت (ب) و با در نظر گرفتن MAE به عنوان تابع loss، مقادیر metric های MSE و MAE را در هر ایپاک برای داده آموزشی و داده تست در یک نمودار رسم کنید (یک نمودار برای معیار MSE و یک نمودار برای MAE). برای داده تست نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی را رسم نمایید. تعداد ایپاک بهینه در این قسمت را مشخص کنید.

ه) ابتدا روابط ریاضی MSE و MAE را بنویسید و سپس نتایج قسمت ج و د را با هم مقایسه کرده و توضیح دهید.

سوال ۲ MLP (Classification)

یک طبقه‌بند با استفاده از شبکه MLP برای دسته‌بندی دادگان [sonar](#) طراحی کنید. این مجموعه الگوی برخورد سیگنال سونار به سیلندر در حالت‌های مختلف است و هدف دسته‌بندی جنس سیلندر (آهن یا سنگ بودن) است. توضیحات بیشتر در لینک قرار داده شده است.

الف) مجموعه داده را با استفاده از تقسیم‌بندی مناسب، به دادگان آموزش ارزیابی و تست تقسیم کنید. درباره‌ی روش تقسیم‌بندی انتخابی خود و مزایای آن توضیح دهید. پیش‌پردازش‌های مناسب را برای مجموعه داده در نظر بگیرید. مدل MLP خود را با حداقل ۲ لایه بسازید. معماری شبکه ساخته شده را ذکر کنید.

ب) نمودار تغییرات دقت و خطای مدل را در هر epoch برای دادگان آموزش و ارزیابی نمایش دهید.

ج) مقادیر خطا، دقت و confusion matrix دادگان تست را گزارش کنید.

د) درباره‌ی معیار خطای مورد استفاده در آموزش شبکه و جایگزین‌های موجود و برتری معیار انتخابی بحث کنید.

ه) آیا معیار دقت استفاده شده عیار مناسبی برای توانمندی شبکه در دسته‌بندی این مجموعه داده است؟ مقدار معیارهای منتخب دیگر را به دست آورید.

و) لازم است مراحل بالا را برای مدل با آموزش از روش stochastic mini batch based استفاده کنید. از اندازه‌های ۳۲, ۶۴, ۱۲۸ استفاده کنید. ۳ حالت را به طور اجمالی توضیح دهید و دلیل تفاوت خروجی در حالت‌های مختلف را شرح دهید. همچنین اندازه batch متناسب را اعلام و علت آن را توضیح دهید.

ح) تفاوت epoch و iteration را بیان کنید و مقدار آن را در قسمت قبل نمایش دهید. مقدار بهینه epoch چگونه به دست می‌آید؟ اگر تعداد epoch ها افزایش یابد چه اتفاقی می‌افتد؟

ط) توابع فعال ساز را در لایه‌های ماقبل آخر تغییر دهید و نتایج را گزارش کنید. مزایا و معایب توابع فعال ساز انتخابی را نسبت به یکدیگر بیان کنید. (از توابع tanh, sigmoid, ReLU استفاده کنید).

ی) افزودن لایه به شبکه تأثیری در خروجی دارد؟ فرضیه خود را با افزودن تعداد لایه‌های مختلف بررسی کنید.

ک) بهترین شبکه به دست آمده با استفاده از تغییر هایپرپارامترهای مربوطه در بخش‌های قبلی به دست آمد. این پارامترها را گزارش کنید. به نظر شما، آیا می‌توان شبکه را بهبود بخشید؟

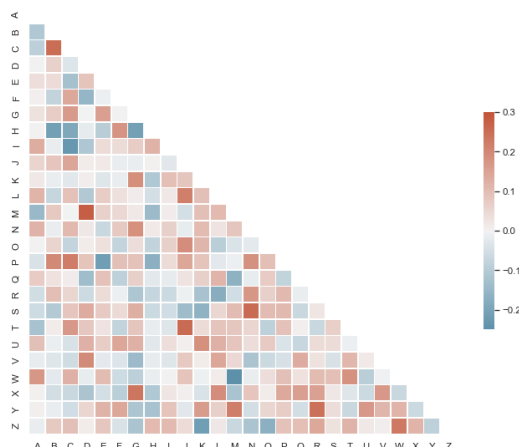
ل) آیا شبکه با نوروں و لایه‌ی کمتر، در epoch های کمتری overfit می‌شود؟ فرضیه‌ی خود را آزمایش کنید و علت را توضیح دهید.

سوال ۳ Dimension Reduction

الف) ماتریس همبستگی^۱ داده سوال یک (شامل ویژگی‌ها و قیمت) را رسم کرده و تحلیل مختصری درباره آن بنویسید.

نکته: از پالت رنگی مناسب همانند شکل زیر استفاده کنید.

نکته: پیشنهاد می‌شود از کتابخانه seaborn.heatmap برای این کار استفاده کنید.



خروجی نمونه

امتیازی - دو سوال زیر را با استفاده از مدل و داده سوال یک حل کنید.

ب) با استفاده از مدل‌های Linear Regression^۲ و Decision Tree (CART)^۳ اهمیت هر ویژگی را به دست آورده و در یک بارپلات نمایش دهید.

نکته: برای این کار توصیه می‌شود از کتابخانه scikit-learn و متدهای DecisionTreeRegressor^۴ و LinearRegression^۵ آن استفاده کنید.

^۱ Correlation Matrix

^۲ [Linear Regression Feature Importance](#)

^۳ [Decision Tree Feature Importance](#)

^۴ [Decision Tree Regressor](#)

^۵ [Linear Regression](#)

ج) در این قسمت با استفاده از روش ^۱Backward Elimination ویژگی‌های مهم‌تر را پیدا کنید. در هر مرحله ویژگی حذف شده را گزارش کنید. زمان انجام این کار را گزارش کنید. مدل را با داده کاهش بعد یافته آموزش دهید و خطای مدل و مدت زمان آموزش را گزارش کنید.

سه سوال زیر را با استفاده از مدل و داده سوال دو حل کنید.

نکته: در دو سوال زیر ابعاد داده را به یک اندازه کاهش دهید تا قابل مقایسه باشند.

د) در این قسمت با استفاده از متد PCA ابعاد داده را کاهش داده و با استفاده از بهترین پارامترهای به‌دست‌آمده برای شبکه سوال دو، این مدل را با داده به‌دست‌آمده آموزش دهید. زمان اجرای PCA و آموزش شبکه را گزارش کنید. دقت و خطای مدل بر روی داده تست را گزارش کنید.

ه) در این قسمت با استفاده از یک شبکه اتوانکودر ابعاد داده را کاهش داده و با استفاده از بهترین پارامترهای به‌دست‌آمده برای شبکه سوال دو، این مدل را با داده انکد شده توسط اتوانکدر آموزش دهید. زمان آموزش اتوانکدر و شبکه را گزارش کنید. دقت و خطای مدل را بر روی داده تست گزارش کنید.

و) جدول زیر را پر کنید. این سه روش را با هم مقایسه کنید.

مقایسه دقت شبکه‌های مختلف

زمان	خطای داده تست	دقت داده تست	
			بهترین شبکه سوال ۲
			AutoEncoder
			PCA

نکته: برای دو روش Autoencoder و PCA مجموع زمان‌های خواسته شده در سوال‌های مربوطه را در قسمت زمان بنویسید.

^۱ [Backward Elimination](#)

نکات:

- مهلت تحویل این تمرین ۲۹ فروردین است.
- گزارش را در قالب تهیه شده که روی صفحه درس در Elearn بارگذاری شده، بنویسید.
- گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است. لطفاً تمامی نکات و فرض‌هایی که برای پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید را در گزارش ذکر کنید.
- در گزارش خود برای تصاویر زیرنویس و برای جداول هم بالانویس اضافه کنید.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست. اما باید نتایج به دست آمده را گزارش و تحلیل کنید.
- دستیاران آموزشی ملزم به اجرا کردن کدهای شما نیستند. بنابراین هرگونه نتیجه و یا تحلیلی که در شرح سوال از شما خواسته شده است را به طور واضح و کامل در گزارش بیاورید. در صورت عدم رعایت این مورد، بدیهی است که از نمره تمرین کسر می‌شود.
- در صورت مشاهده تقلب امتیاز تمامی افراد شرکت‌کننده در آن، ۱۰۰- لحاظ می‌شود.
- برای انجام تمرین‌ها و مینی پروژه‌ها، تنها زبان برنامه نویسی مجاز Python است.
- استفاده از کدهای آماده برای تمرین‌ها به هیچ وجه مجاز نیست. اما برای مینی پروژه‌ها فقط برای قسمت‌هایی از کد و به عنوان راهنمایی برای پیاده‌سازی، می‌توانید از کدهای آماده استفاده کنید.
- نحوه محاسبه تاخیر به این شکل است: مهلت ارسال بدون جریمه تا تاریخ اعلام شده می‌باشد پس از آن به مدت هفت روز بارگذاری ممکن است و به ازای هر روز ۵ درصد نمره کسر می‌شود و در نهایت، پس از بازه تأخیر نمره تکلیف صفر خواهد شد.
- لطفاً گزارش، فایل کدها و سایر ضوابط مورد نیاز را با فرمت زیر در سامانه مدیریت دروس بارگذاری نمایید.

HW2_[Lastname]_[StudentNumber].zip

- در صورت وجود هرگونه ابهام یا مشکل می‌توانید از طریق رایانامه‌های زیر با دستیاران آموزشی مربوطه خانم زهرا دهقانی‌تفتی و خانم نسترن درجانی و آقای علیرضا نیل‌گران در تماس باشید:

zahra.dehghani74@ut.ac.ir (سوال ۱)

nastaran14darjani@gmail.com (سوال ۲)

alirezani14garan@gmail.com (سوال ۳)