

# گزارش کار تمرین پنجم درس یادگیری ماشین

محمد لشکری ۴۰۰۱۱۲۰۸۷

text

## ۱ پیش بینی بارانی بودن روز آتی

تمام ویژگی‌ها به جز تاریخ و وضعیت هوای فردا به عنوان  $X$  و ویژگی وضعیت هوای فردا به عنوان  $y$  انتخاب شده است. دادگان با نسبت  $۰/۲$  به آموزش و تست تقسیم شده اند. اعتبارسنجی به کمک Cross Validation انجام می‌شود.

### ۱.۱ پیش پردازش داده‌ها

ویژگی‌های کیفی با استفاده از LabelEncoder به داده‌های عددی گسسته تغییر یافته اند. داده‌های گمشده در ویژگی‌های عددی پیوسته با استراتژی جایگزینی با میانگین و ویژگی‌های عددی گسسته با استراتژی بیشترین فروانی مدیریت شدند. ویژگی‌های عددی گسسته با StandardScaler نرمال شده‌اند. در نهایت OneHotEncoder روی ویژگی‌های عددی گسسته اعمال شده است.

### ۲.۱ مدل سازی و نتایج

با استفاده از جستجوی شبکه‌ای<sup>۱</sup> برای هایپرپارامترهای «تعداد لایه‌ها» و «نرخ یادگیری» که مقادیر آنها (6,6) و برای نرخ یادگیری 0.0001, 0.001, 0.01 است با معیار میانگین ماکرو صحت<sup>۲</sup>، بهترین نتیجه برای دو لایه با ۶ نورون در هر لایه و نرخ یادگیری 0.01 به دست آمد. نتایج حاصل شده در جدول ۱ قابل مشاهده است. نتایج صحت اعتبارسنجی و دقت اعتبارسنجی به کمک Cross Validation و با ۵ فلد به دست آمده‌اند.

دقت تست	دقت اعتبارسنجی	صحت اعتبارسنجی	صحت کلاس صفر	صحت کلاس یک
۰/۸۶	۰/۸۶	۰/۸۱	۰/۸۸	۰/۷۴

جدول ۱: نتایج پرسپترون دو لایه با ۶ نورون در هر لایه

<sup>۱</sup>Grid Search

<sup>۲</sup>precision

دقت اعتبارسنجی و تست برابر هستند که دلالت بر همگرایی مدل دارد. صحت دادگان کلاس ۱ مطلوب نیست اما صحت داده‌های اعتبارسنجی قابل قبول است. به طور کلی اگر هوا در روز آینده بارانی باشد و ما اعلام کنیم بارانی نیست ضرر بیشتری را به شهروندان وارد می‌کنیم. بنابراین صحت دادگان کلاس ۰ مهم تر از کلاس ۱ است و مدل عملکرد مطلوبی دارد.

## ۲ پیش بینی دما به وسیله شبکه LSTM

دادگان به صورت مقادیر مشاهده شده از یک سری زمانی یک متغیره در نظر گرفته شد و دادگان ۵ روز گذشته برای پیش بینی دمای روز ششم استفاده شد. مدلی که برای حل این مسئله در نظر گرفته شده است یک لایه LSTM با ۶۴ نورون و یک لایه چگال ۳ با یک نورون دارد. دادگان با نسبت ۲/۰ به آموزش و تست و سپس دادگان آموزش با نسبت ۵/۰ به آموزش و اعتبارسنجی تقسیم شدند. مدل با ۱۰۰ تکرار ۴ و اندازه دسته ۵ ۶۴ و تابع هزینه میانگین خطای مربعات، آموزش داده شد. خطای آموزش و اعتبارسنجی در طی آموزش با سرعت خوبی کاهش یافتند. مقدار خطای اعتبار سنجی به خطای تست نزدیک است و مقادیر این دو اختلاف کمی با خطای آموزش دارد که نشان می‌دهد مدل همگرا شده است. در سطر دوم جدول نتایج مدلی را مشاهده می‌کنید که از ۳۰ روز گذشته برای پیش بینی استفاده می‌کند و در آن خطای تست از مدل اول بیش‌تر است پس مدل اول عملکرد بهتری داشته است که دلیل این امر می‌تواند عدم وابستگی دمای یک روز به روزهای دورتر در گذشته باشد.

آموزش به ازای	خطای آموزش	خطای اعتبارسنجی	خطای تست
۵ روز	۸/۰۰	۵/۶۱	۵/۷۳
۳۰ روز	۸/۱۵	۵/۵۹	۷/۱۹

جدول ۲: نتایج مدل LSTM

<sup>3</sup>Dense

<sup>4</sup>epoch

<sup>5</sup>batch size