

## گزارش تمرین دوم مهرانه مقتدائی فر 97222086

### قسمت دوم:

با توجه به داده های داده شد، یک فایل برای عکس های خانه ها و فایل دیگر برای ویژگی خانه ها داده شده است. ابتدا دیتا های موجود در فایل csv را مرتب میکنیم

دو داده ی city و street داده های categorical ما هستند. پلاک street را جدا کردم و در ستون street tag قرار دادم اما مشکلی پیش آمد، چون تمامی street ها با عدد شروع نمی شد و یکی از آنها string اول آن موجود بود با این روش نتوانستم اعمال کنم، چون در scale کردن به مشکل برمیخوردم

کار دیگری که انجام دادم بخش اعشاری bath را از آن جدا کردم و به عنوان ستون half bath قرار دادم سپس داده ها scale کردم.

در مورد داده های عکس نیز تک تک آنها را خواندم و ویژگی های آنها را در آرایه ای ذخیره کردم. به دلیل آنکه عکس ها خیلی بزرگ بودند آنها را به 64\*64 تبدیل کردم.

حال برای ساختن شبکه مورد نظر چون داده های ما دو نوع هستند باید دو مدل شبکه داشته باشیم. اول برای داده های ویژگی خانه یک شبکه ANN و برای داده های عکسی یک شبکه CNN داریم. سپس این دوتا را باهم merge کرده تا شبکه مورد نظر بدست آید.

سپس مدل را compile کرده و از mse به عنوان متریک استفاده کردم (چون مسئله ما رگرسیون است)

### چند نکته :

1. هرچه تعداد نرون های شبکه و تعداد لایه های آن بهتر باشد آنگاه نتیجه نهایی بهتری خواهد بود.
2. خطای نهایی در این مدل توسط MSE محاسبه میشود.
3. به دلیل آنکه عکس های ما کوچکتر شده اند امکان آنکه برخی ویژگی ها از دست رفته باشد امکان دارد.
4. در خصوص داده های csv اگر street , city لحاظ میشد شاید تاثیر بهتری داشت
5. به علت پراکندگی داده های ما و تمیز نبودن آنها شبکه ممکن است اشتباه کار انجام داده باشد

## محاسبه میانگین خطا:

برای محاسبه میانگین خطا ابتدا خروجی که شبکه برای هر مورد پیشبینی میکند را جدا کردم و سپس این خروجی را از قیمت واقعی کم میکنم (قدر مطلق آن را نیز محاسبه می کنیم چون نمی دانیم کدام قیمت بیشتر بوده است) در نهایت برای پیدا کردن میانگین خطاها آن ها را باهم جمع زده و تقسیم بر کل داده های تست میکنیم. مشاهده میکنیم که میانگین خطای بدست آمده بالاتر از 200000 است که این نتیجه خوبی نیست!

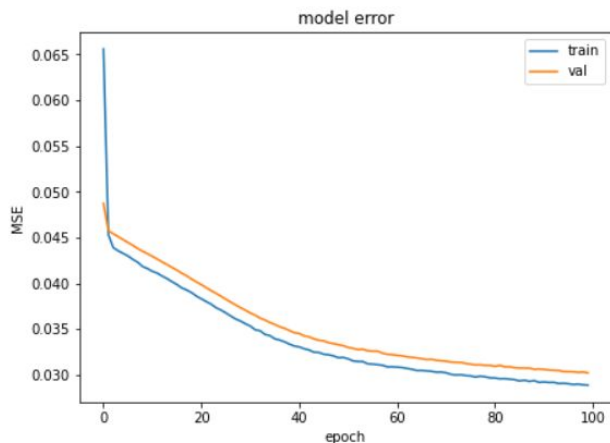
در نهایت یک بار نیز بدون عکس ها شبکه را فقط با داده های CSV آموزش دادم و مشاهده میشود که نتیجه نهایی یعنی میانگین خطا در این مورد کمتر است و این شاید به این معنا باشد که داده های تصویری کمک زیادی به ما نمیکند.

## نتیجه گیری نهایی:

مدل من در کل مدل خوبی نبود چون قیمت را بسیار با اختلاف تخمین میزند و این میتواند به دلیل کم بودن لایه ها و یا نامنظم بودن دیتا ها باشد.

تصویر های ما خیلی کمکی در تخمین زدن قیمت به ما نمیکند چون اگر شبکه را بدون آن train کنیم نتیجه بهتری خواهیم گرفت!

ANN ONLY(without images data)



ANN & CNN Merged

