ایده کلی این بود که ابتدا دو نتورک هر کدام برای پیشبینی قیمت خانه ها بطور جداگانه یکی را فقط با دیتا های عددی و دیگری را فقط با عکس ها ترین کنیم

سپس بعد از آن چند راه مختلف را امتحان کنیم برای تلفیق این دو نوع داده تا به یک شبکه قوی تر برسیم

ابتدا یک نتورک فقط با استفاده از دیتا های عددی میسازیم :

نمای اولیه از دیتا ها نشان میدهد که دیتا های ما شامل 1 ستون برای تعداد سرویس های بهداشتی (که بطور اعشاری نمایانگر نوعی سرویس مستر هم بود که باید جدا شود و در ستونی جداگانه قرار گیرد)

1 ستون برای تعداد اتاق خواب ها

1 ستون برای ادرس خانه (درون شهری)

1 ستون برای شهر ها (اسم)

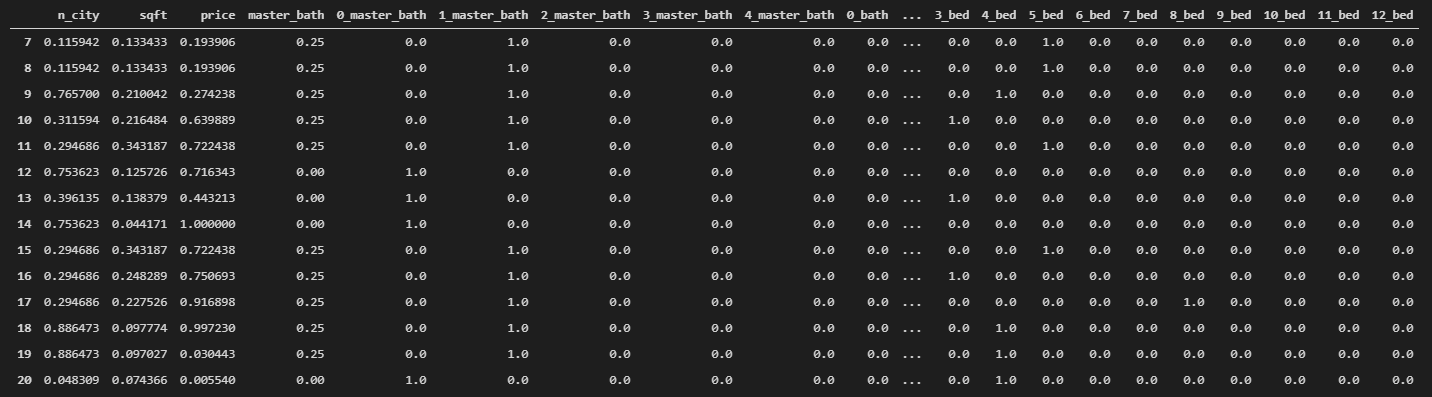
و 1 ستون برای آیدی هر شهر است

ابتدا سعی کردم از ستون شهر و ادرس اطلاعاتی رو استخراج کنم اما اطلاعات انچنان مفیدی رو ارایه ندادند و در جواب نتورک تاثیری نداشت

منتها با جدا کردن ستون سرویس بهداشتی از سرویس بهداشتی مستر و ایجاد یک ستون برای هر مقدار در ستون های اتاق خواب و سرویس بهداشتی و سرویس بهداشتی مستر و ایدی شهر ها نمای کلی به شکل زیر بود

تغییرات بالا در دو مرحله رخ گرفت :

اول فقط از ستون های اتاق خوب و سرویس بهداشتی و سرویس بهداشتی مستر استفاده شد که نتایج ان به شکل زیر است



دیتا های بالا در چهار حالت مختلف امتحان شد (بهینه ساز های Adam و RmsProp همچنین تابع خطا های mae و mse)



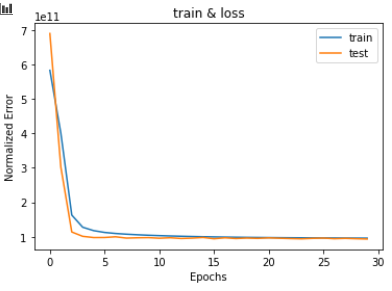
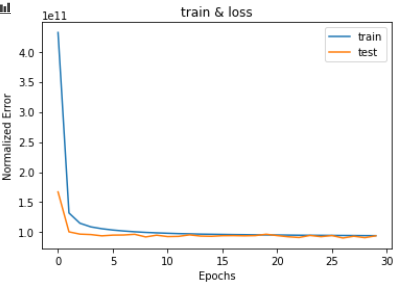
Opt : RmsProp :: Loss : Mean Absolute Error

Opt : Adam :: Loss : Mean Squared Error



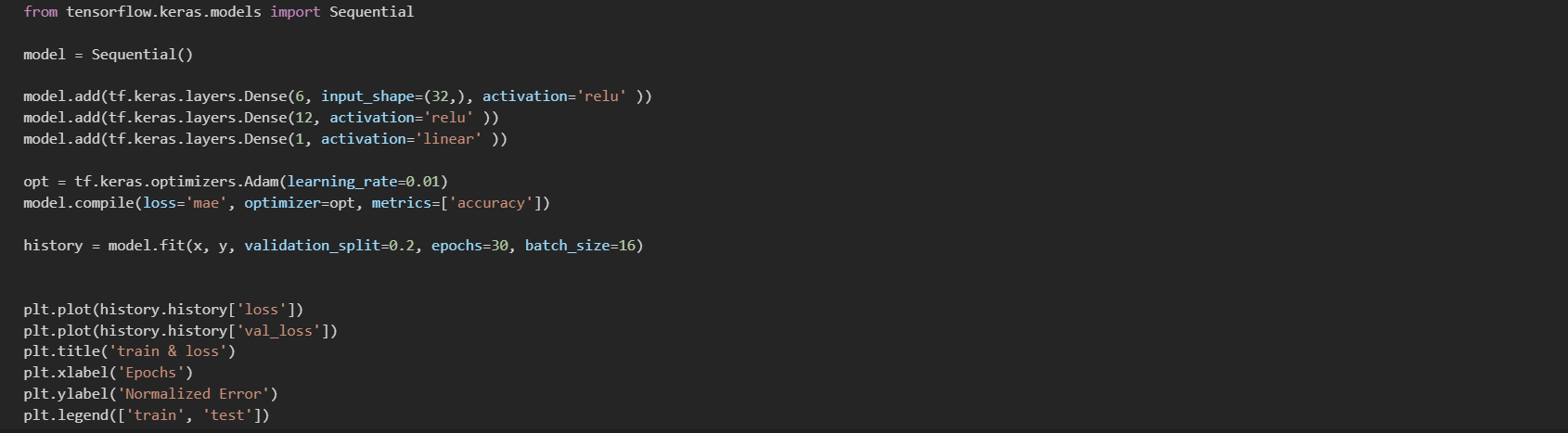
Opt : Adam :: Loss : Mean Absolute Error

Opt : RmsProp :: Loss : Mean Squared Error

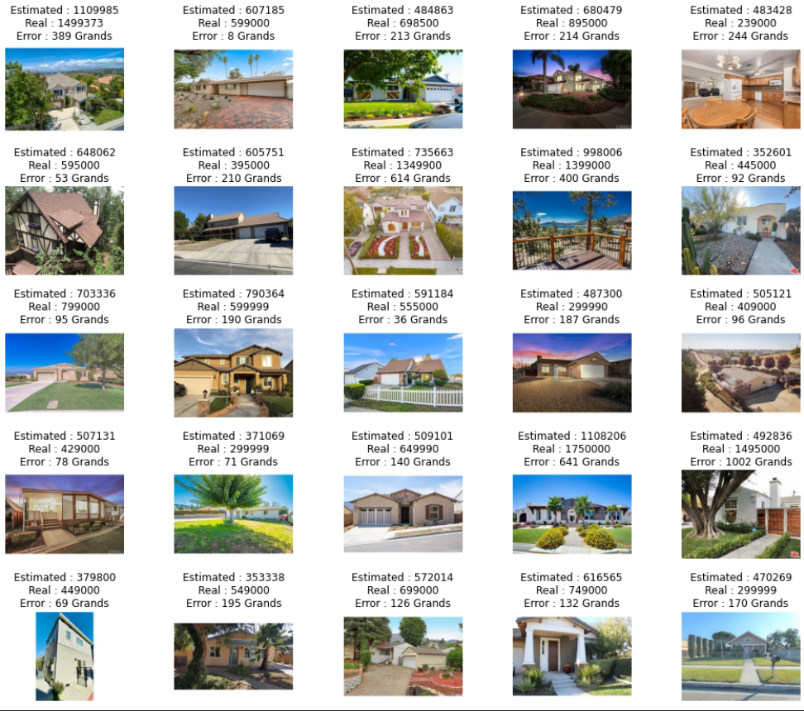
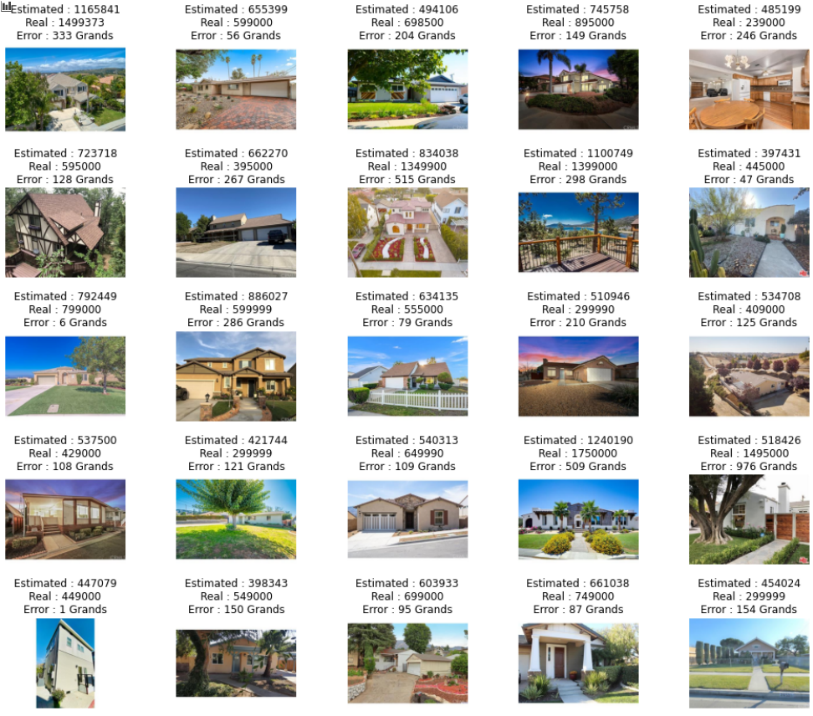


RmsProp

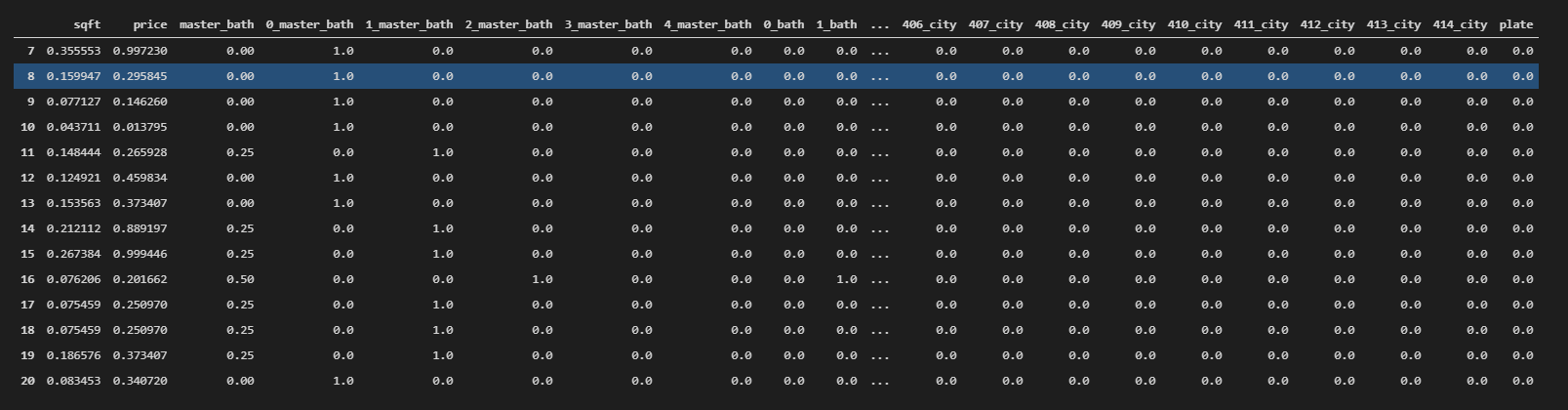
Adam



ساختار شبکه و هایپر پارامتر های استفاده شده در شکل بالا امده است

 در زیر دو عکس را مشاهده میکنیم که سمت راست چند نمونه پیشبینی شده با Adam و mae و سمت چپ همان نمونه ها با Adamو mse

در حالت بعدی آیدی شهر ها را نیز وارد دیتا های ورودی برای شبکه میکنیم نمونه دیتا در زیر امده است



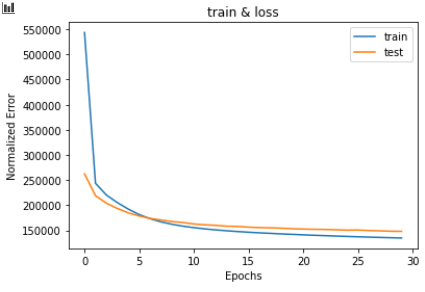
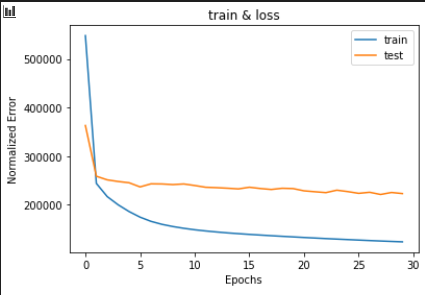
در این حالت 447 ستون داریم در حالی که حالت پیش حدود 30 ستون داشتیم !

در حالت اول (بدون ایدی شهر ها) میانگین خطا روی داده های ترین برابر 221479 و روی داده های ولیدیت برابر 236307 بود

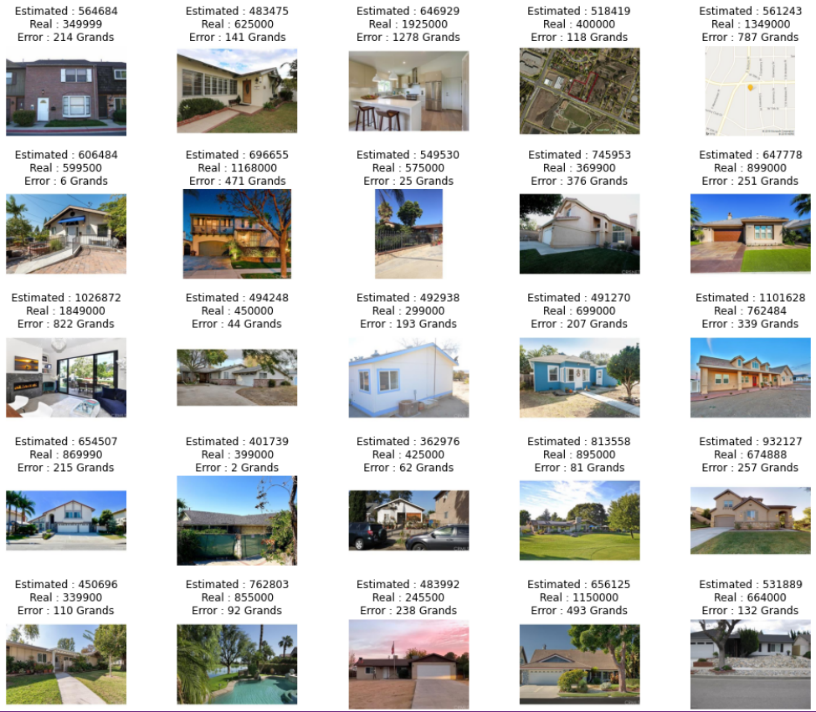
وقتی آیدی شهر ها را نیز وارد کردیم میانگین خطا روی داده های ترین برابر 12753 و روی داده های ولیدیت 2203055 یعنی به اورفیتینگ برخورد کریم

با اعمال دراپ اوت با احتمال 50 درصد روی لایه های Dense و 25 درصد روی لایه های Convolution نتیجه زیر بدست میاید

میانگین خطا روی داده های ترین برابر 138387 و روی داده های ولیدیت برابر 150540

که میتوان ادعا کرد یک پیشرفت است

دو عکس زیر قبل به ترتیب قبل از دخیل کردن آیدی شهر ها و بعد از دخیل کرن آنهاست (سمت چپ : بعد و راست : قبل)

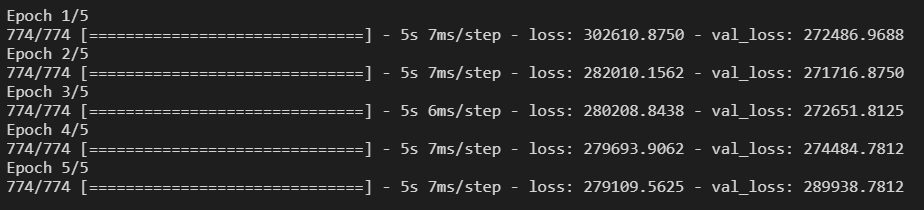


حال نوبت آن است که یک نتورک برای ترین کردن با عکس ها بسازیم



تصویر بالا ساختار نتورک را نشان میدهد برای فیت کردن عکس ها هیچگونه آگمنتیشنی روی آنها رخ نداده است زیرا احتمال اثر منفی داشت

نتایج اموزش شبکه با ساختار بالا برابر است با :



میبینیم که میانگین خطا بسیار بالا تر از شبکه قبلی است و همینطور هم انتظار میرفت زیرا تعدادی از عکس ها هیچ ربطی به قیمت خانه ندارند و همچنین در تعداد زیادی از عکس ها فیچر مشخصی برای استخراج وجود ندارد که بتواند نتیجه مثبتی روی دقت شبکه بگذارد

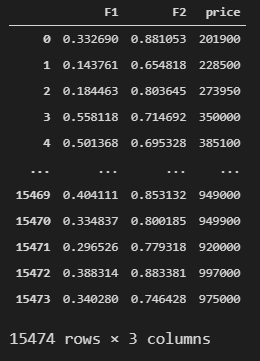
قبل از تلفیق دو شبکه بالا نگاهی به دقت جداگانه آنها می اندازیم که در زیر امده است :





البته ناگفته نماند که با وجود خطای (نه چندان معقول اما قابل قبولی) که از دو شبکه گرفتیم دقت را بسیار کم نشان میدهد که میتواند بخاطر نوع دیتا هایمان باشد در هر صورت خود پیشبینی ها هستند که برای ما اهمیت دارند و نه درصد دقت

برای تلفیق دو شبکه دو راه را امتحان کردیم

اولین راه این بود که خروجی پیشبینی شده برای همه دیتا ها را از هر دو شبکه بگیریم و توی یک دیتا فریم به عنوان دو ستون یکی برای پیشبینی های شبکه اول (فقط با دیتا های عددی) و ستون دیگر برای پیشبینی همه داده ها با شبکه دوم(فقط عکس ها) و ستون سوم قیمت اصلی خانه ها

سپس با دیتا فریم بالا یک شبکه عصبی ساده ANN میسازیم و به عنوان ورودی دو ستون اول را فیت میکنیم

نتیجه این راه در زیر امده است :



میبینیم که نه تنها افزایش دقت مورد توجهی نداشت بلکه نسبت به نتورک اول ضعف هم داشتیم احتمالا به این دلیل که پیشبینی های نتورک دوم نباید اهمیت چندانی داشته باشند و در مقایسه مستقیم با پیشبینی های دقیق تر نتورک اول قرار گیرند

حالت بعدی که امتحان میکنیم این است که نتورک ANN جدیدی بسازیم و به عنوان ورودی 448 ستون به آن بدیم

447 ستون اول همان ستون های نتورک اولمان هستند و ستون 488 ام پیشبینی های نتورک دوم از تمامی داده ها

در زیر میبینیم که نتیجه حاصل شده بسیار بهتر از روش اول و با کمی اختلاف بهتر از نتورک با دیتا های تنها است



با این روش به میانگین خطای 106920 روی دیتا های ترین و 133213 روی دیتا های ولیدیت میرسیم که به نسبت 138000 و 150000

بهبود داشته ایم

در زیر نتیجه هر 4 مرحله را میبینیم







