-۱

الف) لزوما درست نمیباشد

دراینجا مسئله bias/variance tradeoff مطرح است ، افزایش پیچیدگی مدل منجر به کاهش خطای اموزش میشود و زمانی که مدل ما بیش از حد پیچیده باشد معضل high variance و overfitting داریم که باعث افزایش خطای ازمون می شود و برعکس زمانی که مدل ما به شدت ساده باشد معضل high bias و underfitting داریم که باعث میشود خطای اموزش و ازمون افزایش یابد زیرا مدل به سختی نظر خود را عوض میکند.

ب) درست

هرچه داده های کمتری داشته باشیم، مدل ما بهتر می تواند استثناها را در مجموعه آموزشی شما به خاطر بسپارد که منجر به دقت بالا در آموزش اما دقت پایین در مجموعه ازمون می شود زیرا مدل ما چیزهایی را که از مجموعه آموزشی کوچک آموخته است تعمیم می دهد.

ج) درست

افزایش پیچیدگی مدل باعث overfitting مدل بر روی داده های اموزش میشود که کاهش خطای آموزش و افزایش خطای آزمون را به همراه دارد

د) MAE برای داده های پرت و نویزی عملکرد بهتری دارد چون از قدر مطلق برای محسابات استفاده میکند . هرچند در پیاده سازی توابع محاسبه هزینه متریک به صورت پیش فرض از RMSE استفاده می شود ولی به دلیل تفسیر ساده تر MAE (easy to interpretation) الگوریتم MAE بهتر است

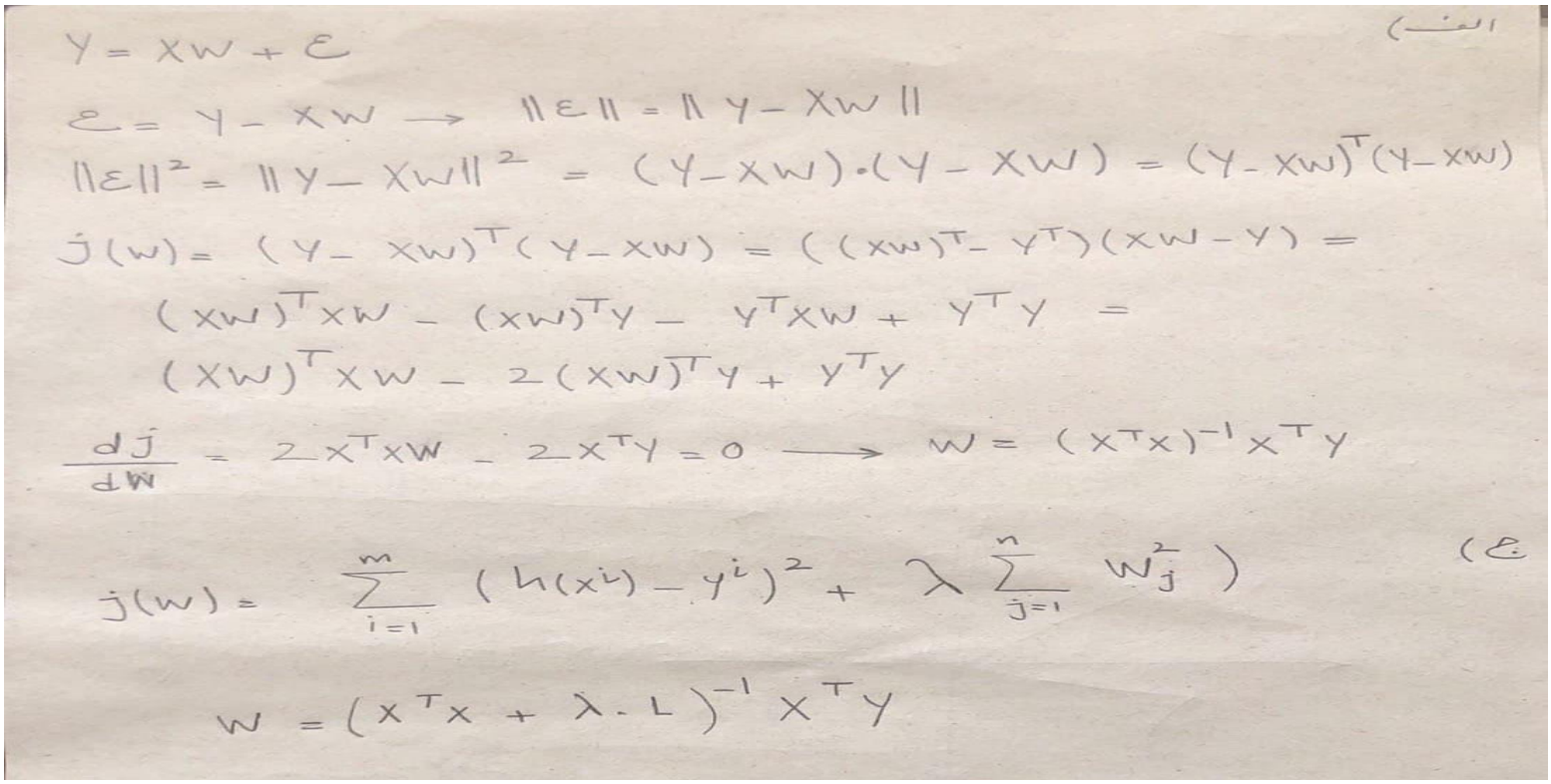
-۲

مدل با پیچیدگی بالا نویز داده های آموزش را بخاطر میسپارد و نمیتواند به مجموعه ارزیابی تعمیم دهد یا زمانی که training برای epochs های زیادی تکرار می شود خطای مجموعه اموزش کاهش می یابد ولی خطای مجموعه ارزیابی شروع به افزایش می کند

- early stopping
- Cross validation
- Regularization

-۳

الف , ج)



ب) زمانی که ماتریس مربعی X^TX وارون نداشته باشد نمیتوانیم از این رابطه به طور مستقیم استفاده کنیم که در دوحالت وارون نداریم :

- ویژگی ها مستقل از هم نباشند و ستون وابسته بهم داشته باشیم
- #m << #n

وقتی ماتریس وارون ندارد میتوان از sodu inverse استفاده کرد ولی جواب تقریبی و غیر منحصربفرد است یا میتوان از regularization استفاده کرد

برای کار با تعداد feature کم گزینه خوبی است اگر تعداد featureها زیاد باشد محاسبات پیچیده میشود

د) چون قصد داریم cost function را minimum کنیم جمله نرمال ساز که اضافه کردیم نیز min میشود و باعث می شود W_j ها که ضرایب چندجمله ایی هایمان هستند نیز کم شوند در نتیجه جلوی زیاد شدن Xها با توان بالا گرفته میشود.

-٤

الف) SGD یا MSGD بهترین گزینه هستند زیرا نیاز به بارگذاری کامل مجموعه اموزش برای انجام یک مرحله از Gradient descent نداریم و normal equation گزینه مناسبی نیست زیرا پیچیدگی محاسباتی در بدست اوردن معکوس ماتریس داریم

ب) روش normal equation نیازمند feature normalizing نیست به همین دلیل برای مجموعه داده با مقیاس های پراکنده عملکرد بهتری دارد ولی در عوض روش gradient descent برای همگرایی سریع تر نیازمند feature scaling می باشد

-٥

افزودن داده اموزشی منجر به کاهش variance می شود. اما تاثیری روی bias ندارد

-٦

خط چین مشکی خطای آموزش

خط چین زرد مربع بایاس

خط قرمز خطای ناشی از نویز

خط چین ابی واریانس

خط سبز خطای آزمون