|  |
| --- |
| katayoun-kobraei  99222084 |

**تمرین اول**

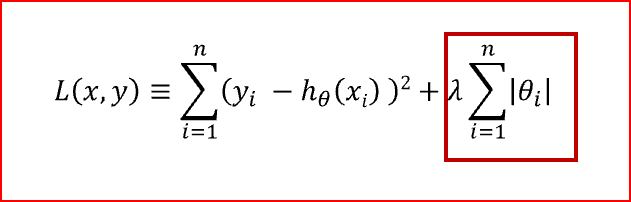
|  |  |
| --- | --- |
|  | l1 And l2 regularization |

این نوع نرم پارامتر وزنی را تا نزدیک صفر کوچک میکنیم. و اگر داده ها نزدیک به صفر یا :L1 regularization

feature

نرمالیزه شده باشند اعمال این نرم باعث کم شدن تعداد های داده در الگوریتم یادگیری میشود. این کار با انتساب

وزن های صفر به ویژگی های کمتر موثر و وزن های بزرگ تر به ویژگی های مهم تر انجام میشود.



|  |  |
| --- | --- |
|  | این نوع نرم وزن های مدل یادگیری را وا میدارد تا کوچک باقی بمانند اما برخلاف نرم :L2 regularization قبلی مقادیر را صفر نمیکند پس منجر به ایجاد یک مدل خلوت نمیشود. این نرم در کار با داده های پرت خوب کار  نمیکند. چون در نقاط پرت خطا پیش بینی مدل زیاد میشود و جمله پنالتی این نرم وزن های مدل کوچک تر خواهند شد.  این مدل زمانی جواب بهتری خواهد داشت که تمام ویژگی های داده ورودی روی مقدار هدف پیش بینی شده ان تاثرگذار باشد و هم چنین وزن های داخل مدل به طور تقریبا ساوی مقداردهی اولیه شده باشند. D:\TERM 5\Nueral network\L2-Regularization.png |

|  |
| --- |
| THE DIFFERENCE BETWEEN L1 AND L2 |

L1

## به جز تفاوت در فرمول این دو نرم تفاوت های دیگری نیز دارند. میدانیم نرم یک مدل خلوت تر ایجاد میکند. همچنین با اثردهی زیرمجموعه های متفاوت از ویژگی ها، مدل های متفاوتی ایجاد میشوند. ولی در نرم تنها یک مدل یادگیری خواهیم داشت چون براساس استفاده از زیرمجموعه های متفاوت از ویژگی ها عمل نمیکند. در نرم ما داریم در حالی که در نرم دیگر نداریم. زمانی که مقدار هدف پیش بینی تابعی از همه ویژگی هاست نرم یادگیری بهتری را سبب میشود. به طور کلی نرم میتواند مدلی ساده و قابل تفسیر ایجاد کند ولی برای الگوریتم های پیچیده عملکرد مناسبی ندارد. برخلاف ان نرم میتواند الگوریتم های پیچیده را در داده ورودی یاد گرفت.

L2

L2

L1

Feature selection

L1

L2

|  |
| --- |
| which one should be prefered? |

## بسته به مسئله ای که با ان مواجه هستیم ترجیح ما فرق خواهد داشت. اگر دنبال راه حلی قوی تر و با امکانات بیشتری باشیم بهتر خواهد بود. این نرم هزینه محاسباتی بیشتری خواهد داشت این درحالی است که نرم هزینه محاسباتی کمتری خواهد داشت.

L2

L1

|  |
| --- |
| describe intuitivly how each efects the model weight |

L1

## در با کم کردن مقدار کمی از وزن در هر تکرار وزن ویژگی های غیر اطلاعاتی را مجبور به صفر میکندو در نتیجه در نهایت وزن صفر میکند. برای مدل اگر پیچیدگی مدل را تابعی از وزن ها در نظر بگیریم، پیچیدگی یک ویژگی متناسب با قدر مطلق وزن ان خواهد بود. در نرم وزن هارا به سمت صفر میبرد، اما انهارا دقیقا صفر نمیکند. جوری عمل میکند که درصد کمی از وزن ها را در هر تکرار حذف کند. بنابر این وزن ها هیچ وقت صفر نخواند شد.

L2

L2

## تمرین دوم:

## در بخش اول تعداد لایه های پنهان را تغییر میدهیم. شاید این کار دقت را افزایش دهد و یا شاید هم نه ولی در نهایت تا حدی جواب خواهد داد و در نهایت باعث خواهد شد.

overfiting

dropout

## در بخش دوم با کردن برخی نورون ها میتوانیم دقت را بالا ببریم. و در نهایت سرعت لرنینگ بالا برود و به نخوریم.

overfiting