قضیه self supervise متفاوته با اینکه میگیم unsupervised سعی میکنیم از خود داده ها label در بیاریم به جای خبره ای که که label میده

مدل های generative میتونه ی prompt بگیره یا هیچی نگیره بعد برامون text و video و.. تولید کنه

تولید تصویر (text to image)

Big gan -🡪 diffusion model

On shot بیا یک مثال بهش نشون بده

Few shot چندتا مثال نشون بده

هر شبکه ای با یک لایه مخفی میتونه هر فانشنی رو تخمین بزنه

اگر شبکه Boolean باشه

خب میشه ترکیبی از and و or ها یعنی جاهایی که یک شده رو نگه میداریم و با ترکیب and اینها ( یعنی x های یک ردیف ) خروجی های را or میکنیم

ولی حداقل تعداد نورون چقدره ؟ اینو از جدول کارنو در میاره

اما بعضی شبکه ها خیلی بزرگ میشن ! قدرت دیپ شدن اینجا معلوم میشه --> رای اکثریت

اگر شبکه classification باشه چی؟

خب مثلا یک فضای محدبه که هر نود میاد یک ضاع رو محدود میکنه و در اخر and اینا

اگر چندتا محدب داشته باشیم هر محدب ی دیپ برا خودش میخواد و or برای تجمیع با ی لایه نشد ک

برای اونجا میایم نقطه نقطه قضیه رو در نظر میگیریم

اگر شبکه پیوسته باشه چی؟

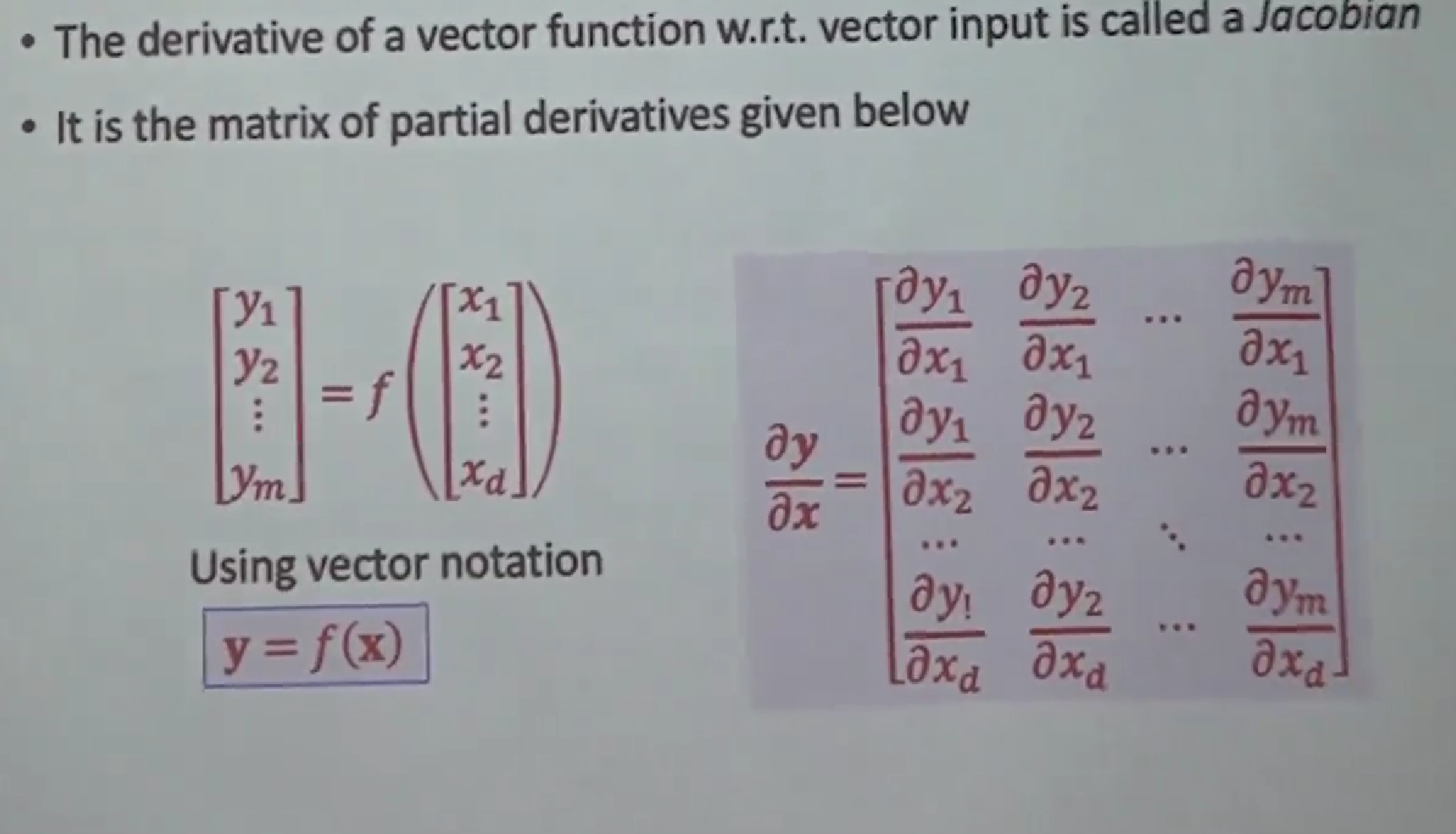
هر کدوم یک یک پالس فانکشن

اوکی deep شدن تعداد ریجن رو خیلی زیاد میکنه اما توابع ما رو ممکنه نتونه تخمین کنه چون ی حال fold داره باید شکل توابع ما رو داشته باشه

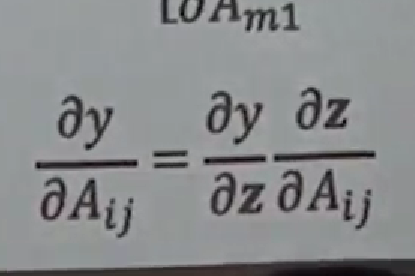
همه اون loss function هایی که قبلا معرفی کردیم مثال multiclass و باینری و reg اینا همه وقتی توزیع احتمالاتی رو بنویسیم و max likelihood رو حساب کنیم بدست میاد

ماتریس جی کوبیان :

یکسری بردار ورودی داریم که تابع f روش اعمال و خروجی حاصل وقتی میایم نسبت به بردار ورودی ازش مشتق میگیریم خروجی میشه ماتریش جی کوبیان



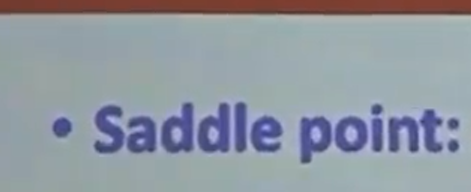
این کجا به کار میاد؟ ما بعدا با تنسور ها رو به رو میشیم وقتی میخوایم خروجی رو مشتق بگیریم نسبت به وزن ها



در واقع در فرایند ما پس یک ماتریس قطری داریم ک میشه مشتق ما نسبت به w

که هر عصر میشه مقدار xj ما

How mini batch work? Todo



جایی که گرادیان صفره اما اگر در یک جهت حرکت گرادیان زیاد یک جهت دیگر گرادیان کم

سوالام ؟  
SDG

How mini batch work

اتا چیه مگ چندتا لرنینگ ریت داریم

چند جمله ای نیوتون

تیلور اکسپرشن

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

مشکلی که regularization حل میکرد این بود که وقتی ما مدل رو بزرگ میکردیم بایاس زیاد میشد و generalization کم میشد و روی داده ترین فقط فیت میشدیم

روش ها :

1.L2 norm

یک جمله توان دو میومد اضافه میکرد به لاس تا اسموس تر کنه

