Session 18:

1404, 01, 31

ابتدا صحبت در مورد اینکه چطور می‌توان از exercise‌ها، رفع پودره تولید کرد.

می‌خواهیم ببینیم از مدل‌های ساده می‌توان برای شبکه‌های عصبی استفاده کرد؟ و چطور Non-linear می‌توان interaction‌های پیچیده‌ی n-body را capture کند

چند نکته: هم برای test و validation ⇐ (I) دقیقاً مانند train نباشد اما خیلی دور هم نباشد

(II) قبلاً ندیده باشیم آن را

learning theory

مدل داده‌ها عالم را ⇐ (D) نداریم ⇐ f ← function‌ها نداریم.

H (Hypothesis class) ← مدل solution ما بهینه است، ← پیش فرض اولیه در ذهن.

Loss function ⇐ تعریف کرده

اگر اختلاف L_S از L_D کمتر از ϵ باشد ⇐ آنگاه مدل خوبی است

در learning ما فرض می‌کنیم data‌ها iid هست

در این مدل‌ها یک trade off بین Bias و complexity وجود دارد.

دلیل اینکه پیش می‌آید (I) overfitting ← بدلیل اینکه بالا بسیار complex می‌شود مدل ما ⇐ بدلیل همین باید پیش فرض‌هایمان را بگذاریم.

(II) با Regular کردن ⇐ complex ↑ آنگاه loss ↓

مدل‌هایی بدون بهتر کردن learning.

ما می‌توانیم مجموعه 20k ورودی را مدل کنیم چون بوشش یا حافظه‌ی ما 2^{20k} زیاد ⇐ با آمدن بدون large language model‌ها، ما زیاد لازم داریم اما sample کردن

مجموعه این‌ها سخت است.