## 李宏毅2023春机器学习课程

P2

①Chatgpt真正做的事：文字接龙——本质上是一个参数量非常大的函数，对于一个句子作为输入，输入后面可能出现的文字的概率，将其中概率最大的字输出，直到输出终止符。

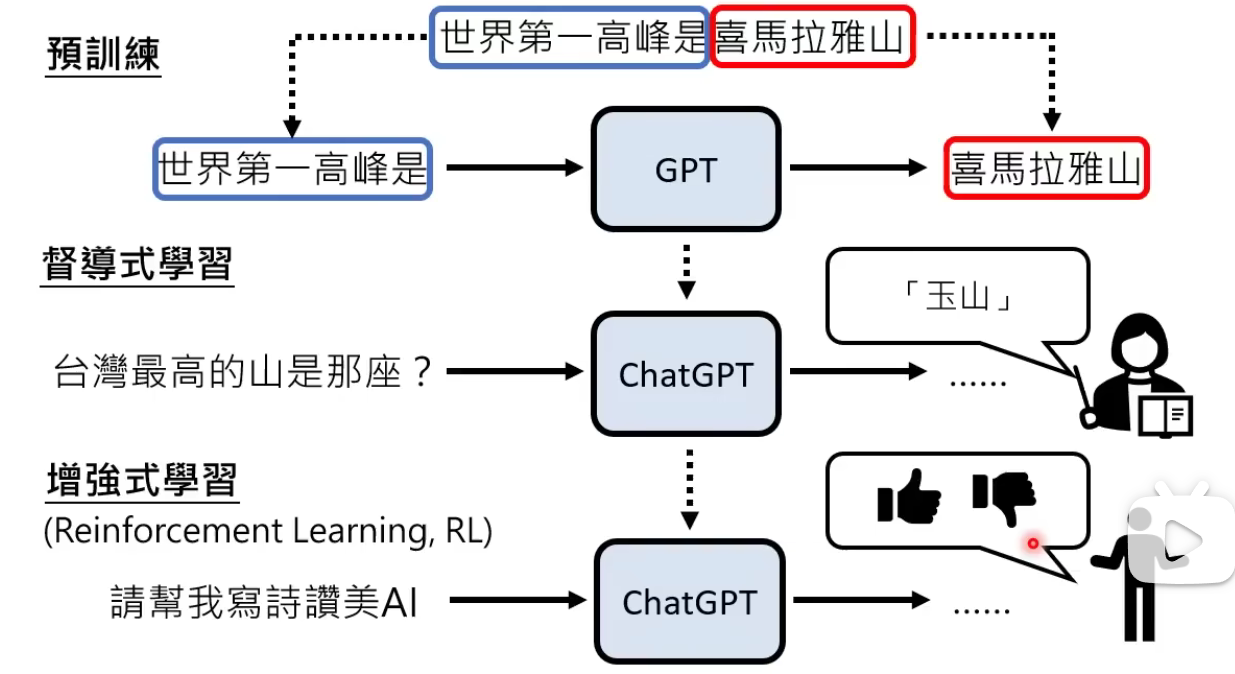
②Chatgpt本身不联网。可能有超过1700亿的参数，在进行训练过程中采用了大量的网络信息和人工输入的知识，但在发布后是不联网的。

P3

①Chatgpt（generative pre-trained transformer）背后的关键技术——预训练

②Chatgpt是GPT经过人类老师提供信息进行监督学习得来的。在Chatgpt3以及之前，都是直接通过网络知识进行的学习，这个过程被称为预训练（往往是自监督学习）。在预训练模型上进行的继续学习称之为微调(finetune)。

③当在预训练模型上，用一种语言训练这种任务，模型很可能会在其它语言上也掌握了相同的任务。



P4

Chatgpt带来的问题：

1.如何精准提需求

2.如何修正它的错误

3.侦测AI生成内容

4.泄密信息

P6

Chatgpt学习的四个阶段：

1.学习文字接龙

2.人类老师引导文字接龙的方向

3.模仿人类老师的喜好

4.用增强学习向模拟老师学习

P7

回归问题

①一般步骤：

1.找一个model

2.定义衡量function好坏的方式（loss function）

3.找到最好的function

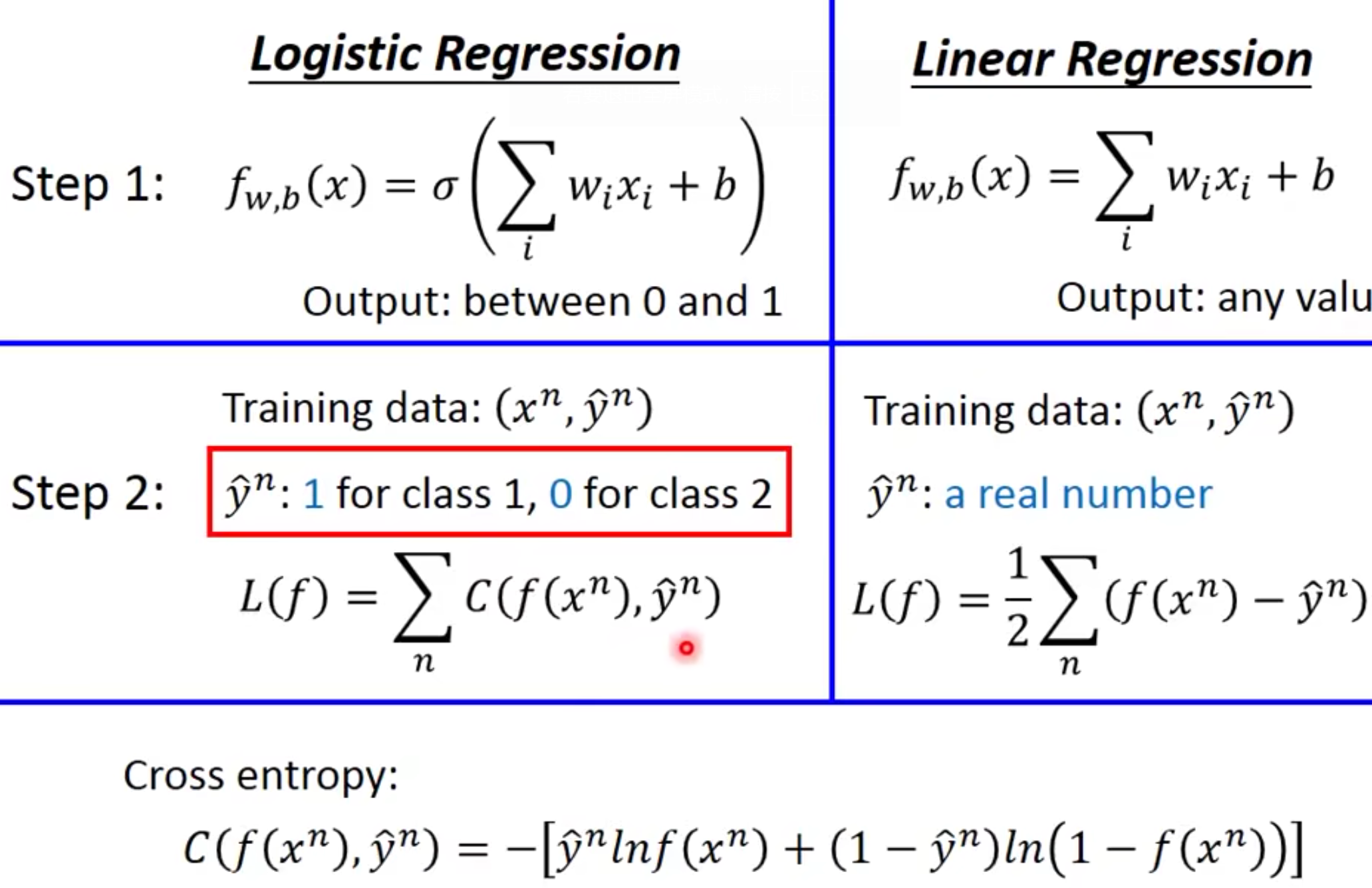
②正则化

加入L2正则化是因为参数值接近0的function比较平滑，对输入数据变化不敏感。这种惩罚迫使模型寻找参数较小的解。参数较小的模型通常更平滑、复杂度较低。

随着λ增大，正则化项的影响力也会增大，这会迫使模型的参数w更接近于零。模型更简单：较大的λ会导致模型更加平滑，复杂度降低，决策边界变得更简单。如果λ过大，模型会变得过于简单，可能无法捕捉到数据的真正结构，这会导致欠拟合，即模型在训练集和测试集上的表现都不好。

P8，9

分类、logist回归（最大似然的思想）



①采用交叉熵而不是mse的原因：均方误差在处理概率输出时，其梯度会较小，尤其是在预测概率接近0或1时。因为逻辑回归的输出是通过 sigmoid 函数得到的，这种非线性映射会导致梯度在远离 0.5 的区域迅速变小，影响训练效率。使用 MSE 可能导致梯度消失。

②判别模型和生成模型：

判别模型的主要任务是直接学习输入数据与目标标签之间的关系，不能生成新的数据点或样本，因为它们没有学习数据的分布。（如神经网络）

生成模型旨在学习输入数据的联合概率分布，然后推导条件概率，可用于生成数据。