

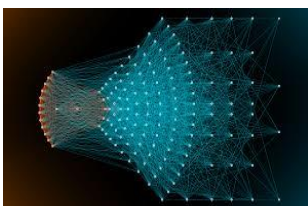
一个最简单的神经网络项目。



李哥考研

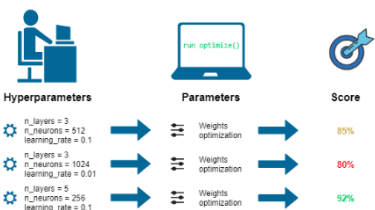


Data: 一般输入是文件地址, 或者数据内容,
输出是一个 存储了数据x,y的数据结构。
torch中一般用dataloader 来装载。



Model: 定义自己的模型

输入x, 输出预测值



hyperPara: 除模型外的超参

一般包含: 学习率, 优化器,
损失函数等

回归损失函数: $LOSS = |\hat{y} - y|$

训
练
流
程



Dataset:

init

getitem

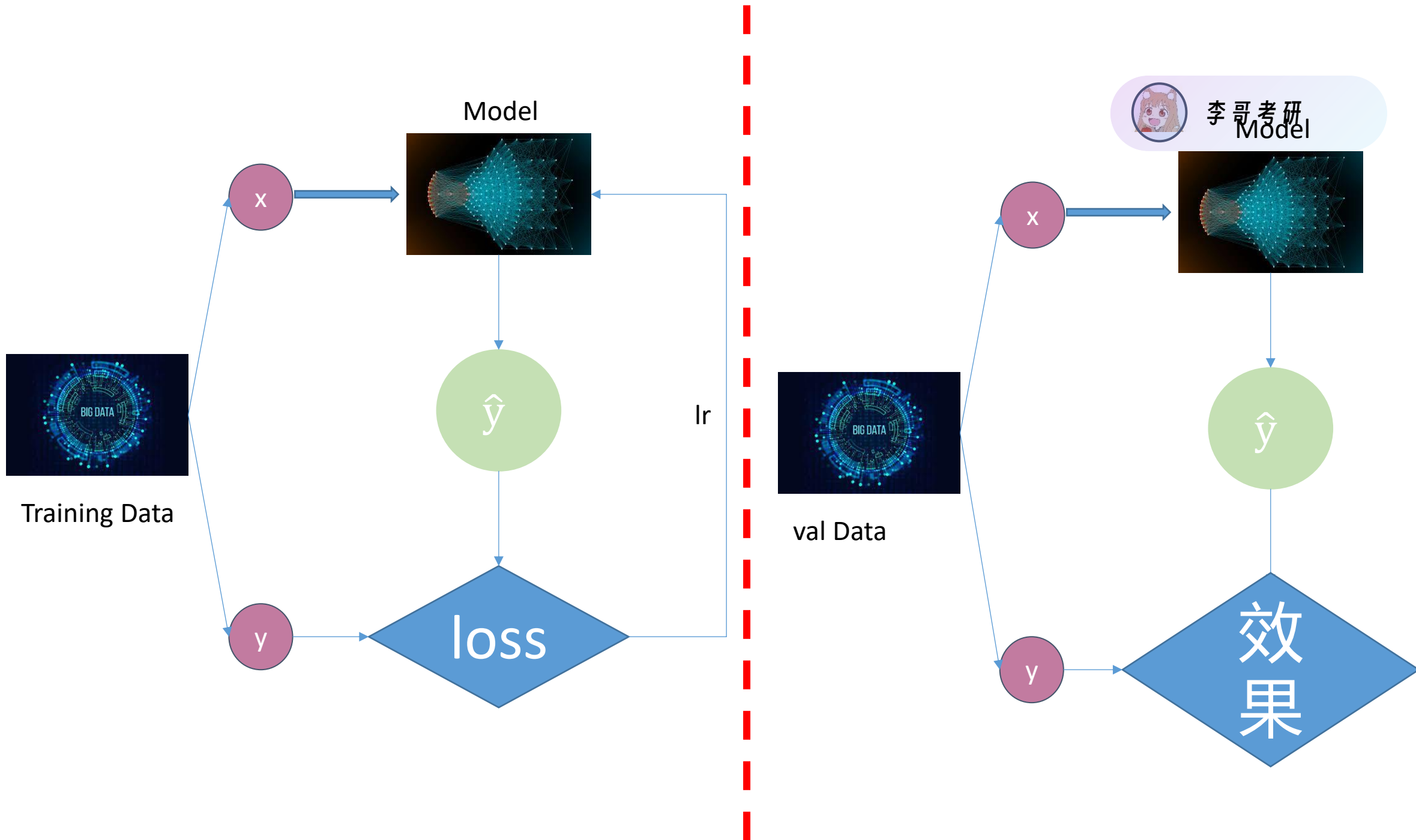
len



Model

init

forward

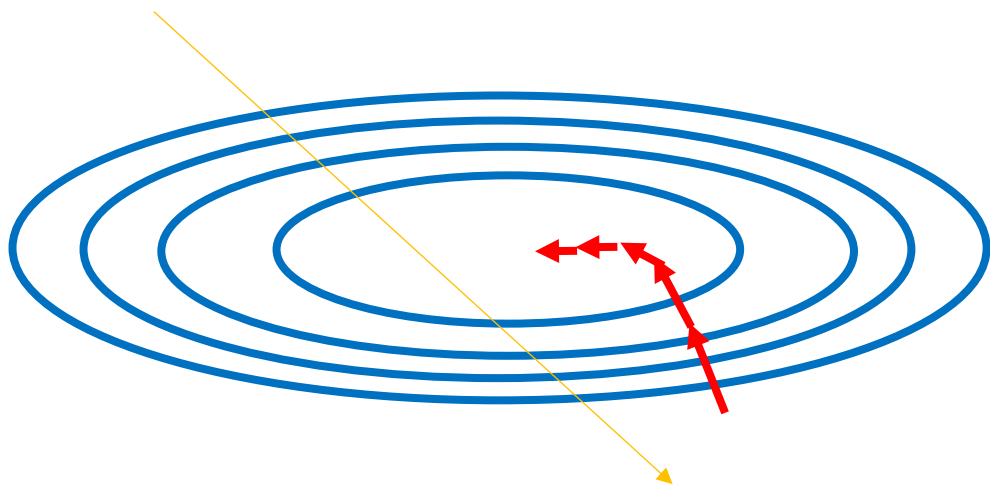


取数据



李哥考研

- 我们一定要用所有的数据算一次loss? 这样会不会有什么问题,

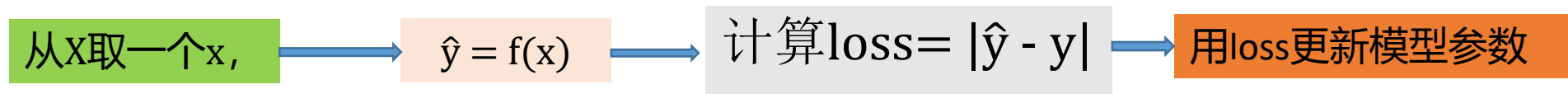


我这一步要走多远?

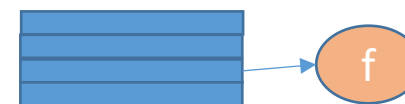
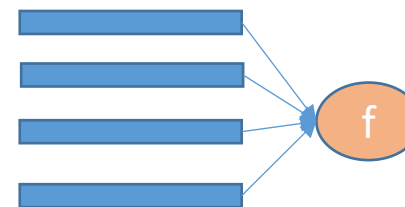
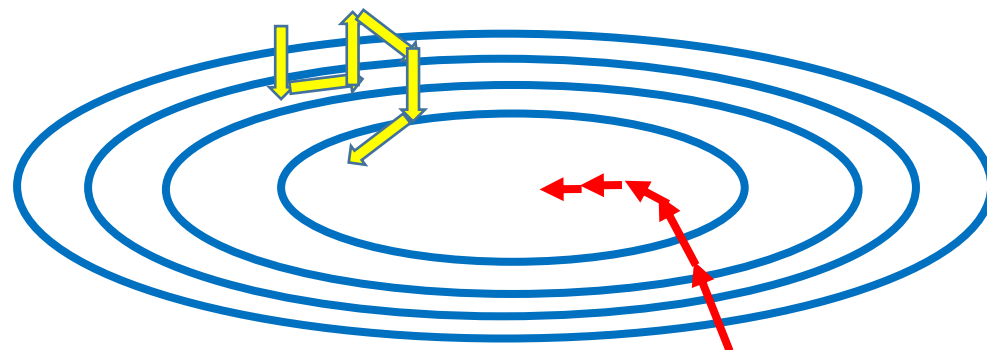


一个数据更新一次就好吗？

loss.backward



一个样本
更新一次
模型



一个batch
更新一次
模型

快!



Optimization of New Model

$$\theta^* = \arg \min_{\theta} L$$

➤ (Randomly) Pick initial values θ^0

➤ Compute gradient $\mathbf{g} = \nabla L^1(\theta^0)$

$$\text{update } \theta^1 \leftarrow \theta^0 - \eta \mathbf{g}$$

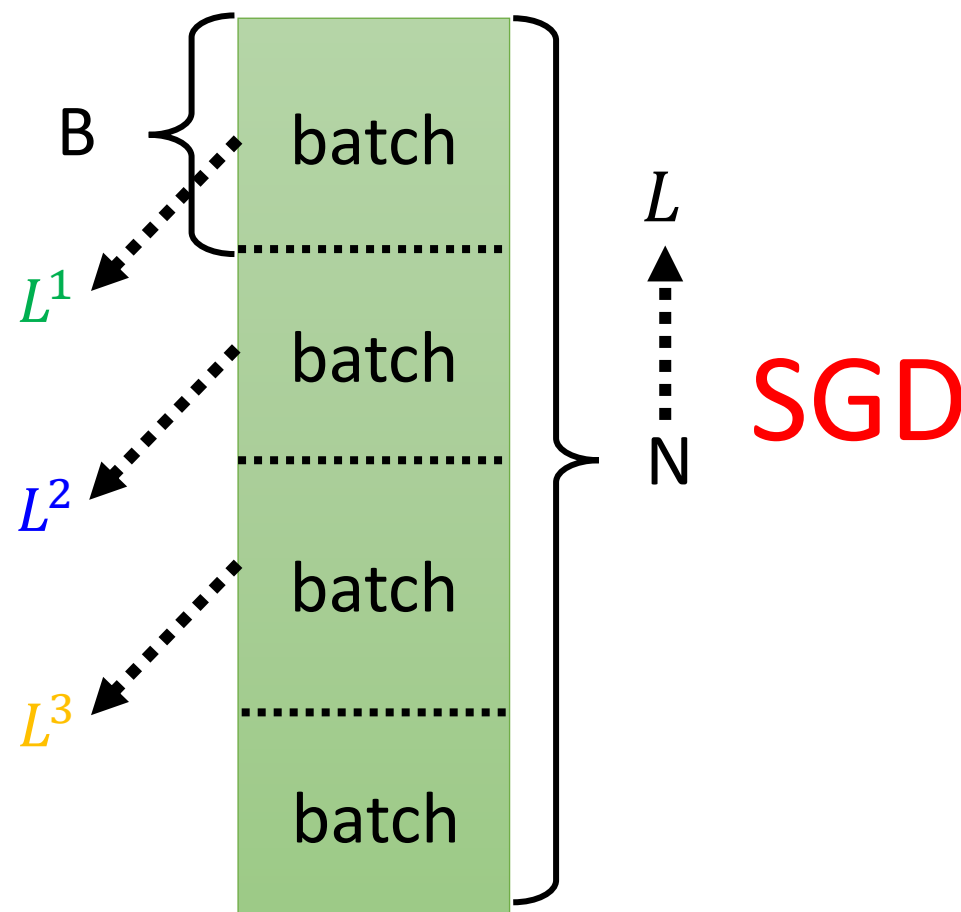
➤ Compute gradient $\mathbf{g} = \nabla L^2(\theta^1)$

$$\text{update } \theta^2 \leftarrow \theta^1 - \eta \mathbf{g}$$

➤ Compute gradient $\mathbf{g} = \nabla L^3(\theta^2)$

$$\text{update } \theta^3 \leftarrow \theta^2 - \eta \mathbf{g}$$

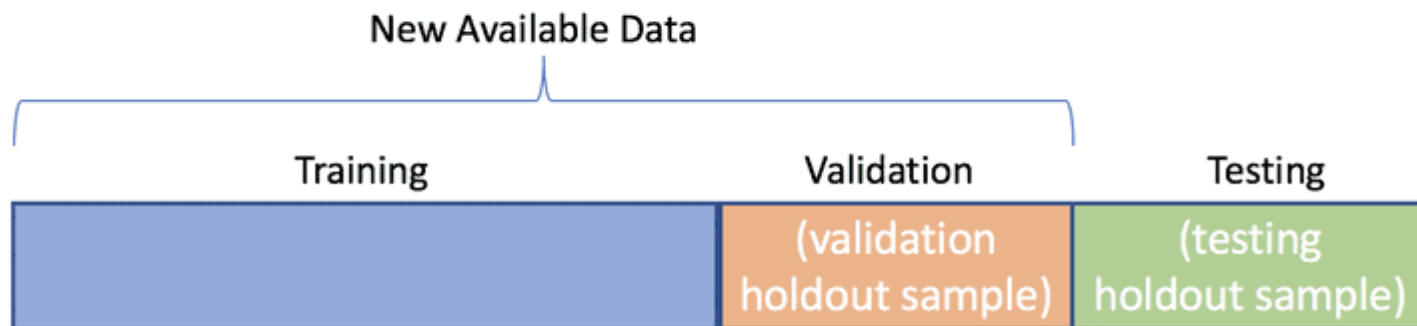
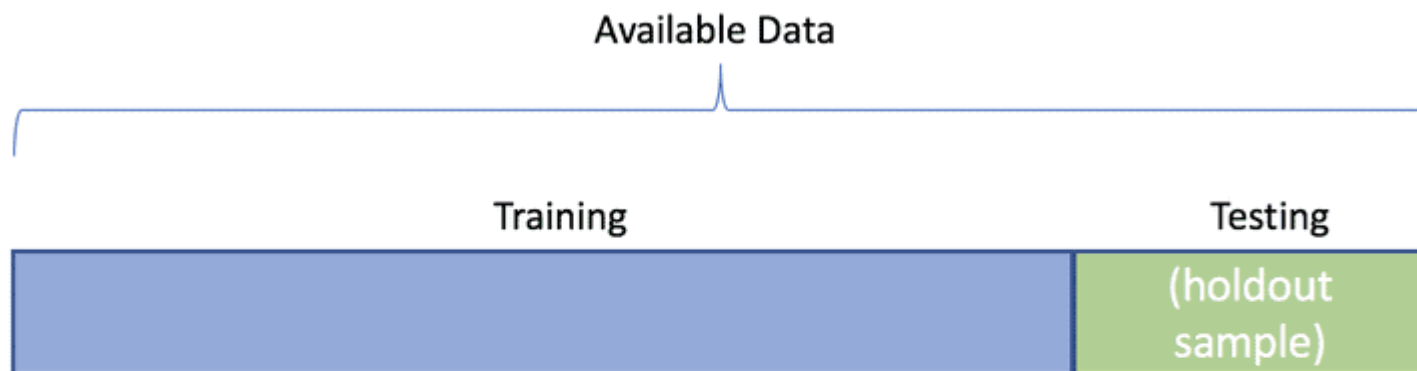
1 **epoch** = see all the batches once



训练集，测试集，和验证集。



李哥考研

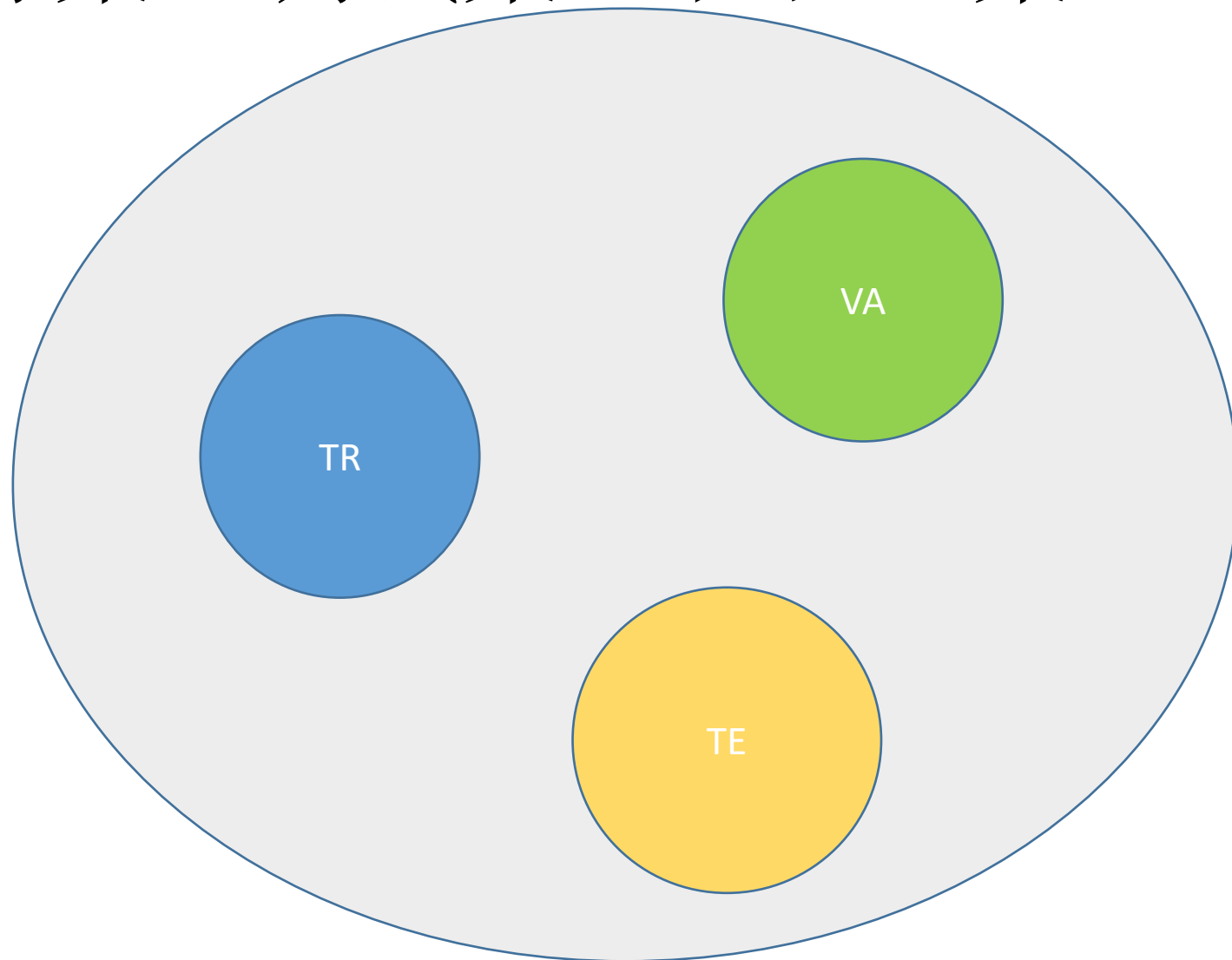


训练集，测试集，和验证集。



李哥考研

要随机



学校/社会

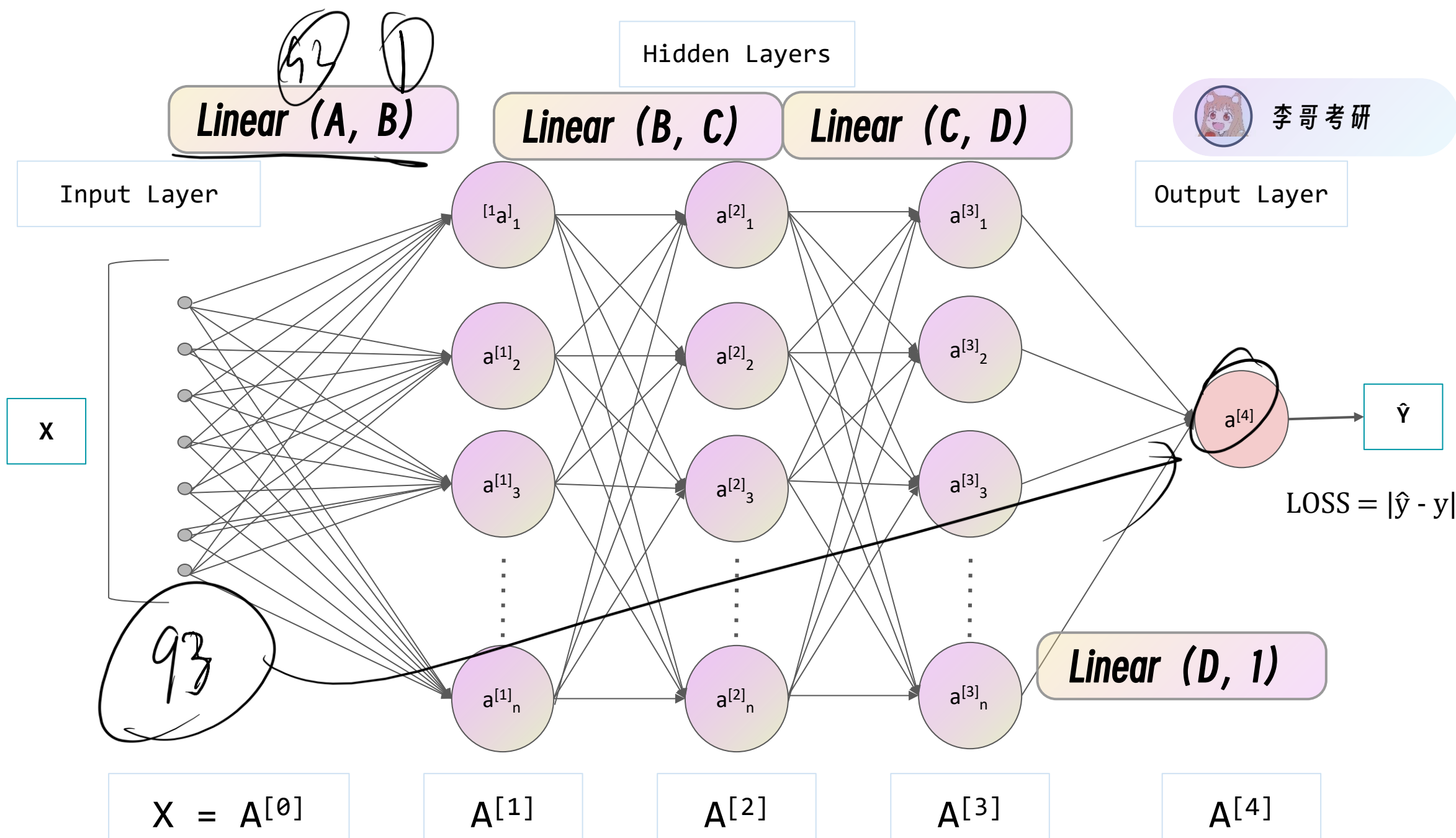
外貌

性格

财富

内涵

恋爱几率



尝试设计网络并计算参数。



李哥考研

Linear (3, 4)

有多少参数量

Linear (4, 1)

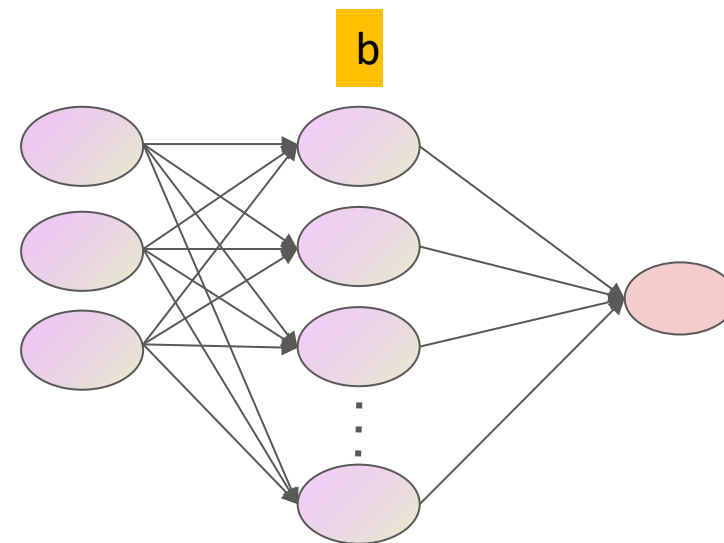
Linear (1000, 1000)

?

$$\text{Num_weights} = 3 \times 4 + 4$$

$$\text{Num_weights} = 4 \times 1 + 1$$

$$\text{Num_weights} = 1000 \times 1000 + 1000$$



新冠病毒感染人数预测。



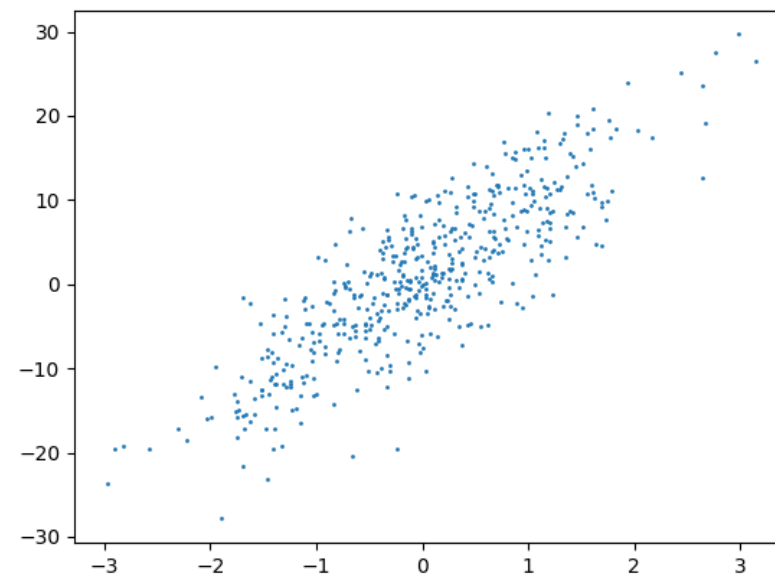
李哥考研

说的是啊 这个美国，有40个州，这四十个州呢，统计了连续三天的新冠阳性人数，和每天的一些社会特征，比如戴口罩情况，居家办公情况等等。现在有一群人比较坏，把第三天的数据遮住了，我们就要用前两天的情况以及第三天的特征，来预测第三天的阳性人数。但幸好的是，我们还是有一些数据可以作为参考的，就是我们的训练集。



新冠病毒感染人数预测。

- 正则化: $\text{loss} = \text{loss} + W * W$
- 相关系数: 线性相关。 (SelectKBest)
- 主成分分析 PCA。





Optimization of New Model

$$\theta^* = \arg \min_{\theta} L$$

➤ (Randomly) Pick initial values θ^0

➤ Compute gradient $\mathbf{g} = \nabla L^1(\theta^0)$

$$\text{update } \theta^1 \leftarrow \theta^0 - \eta \mathbf{g}$$

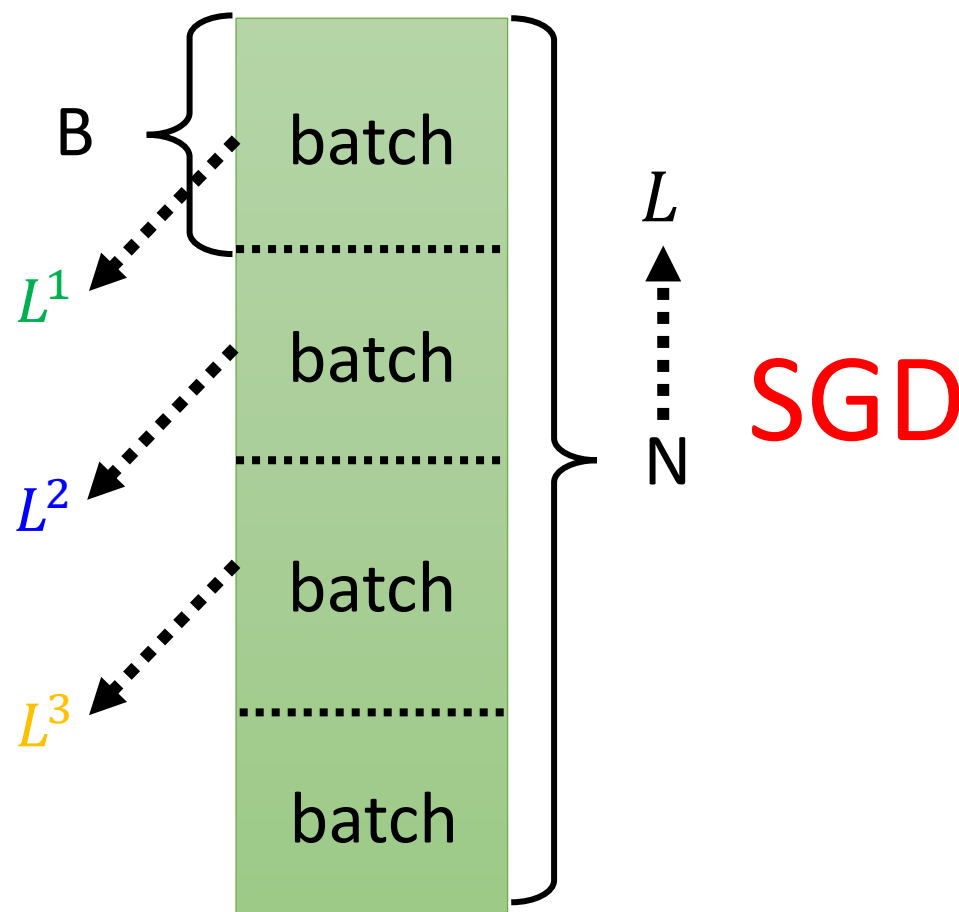
➤ Compute gradient $\mathbf{g} = \nabla L^2(\theta^1)$

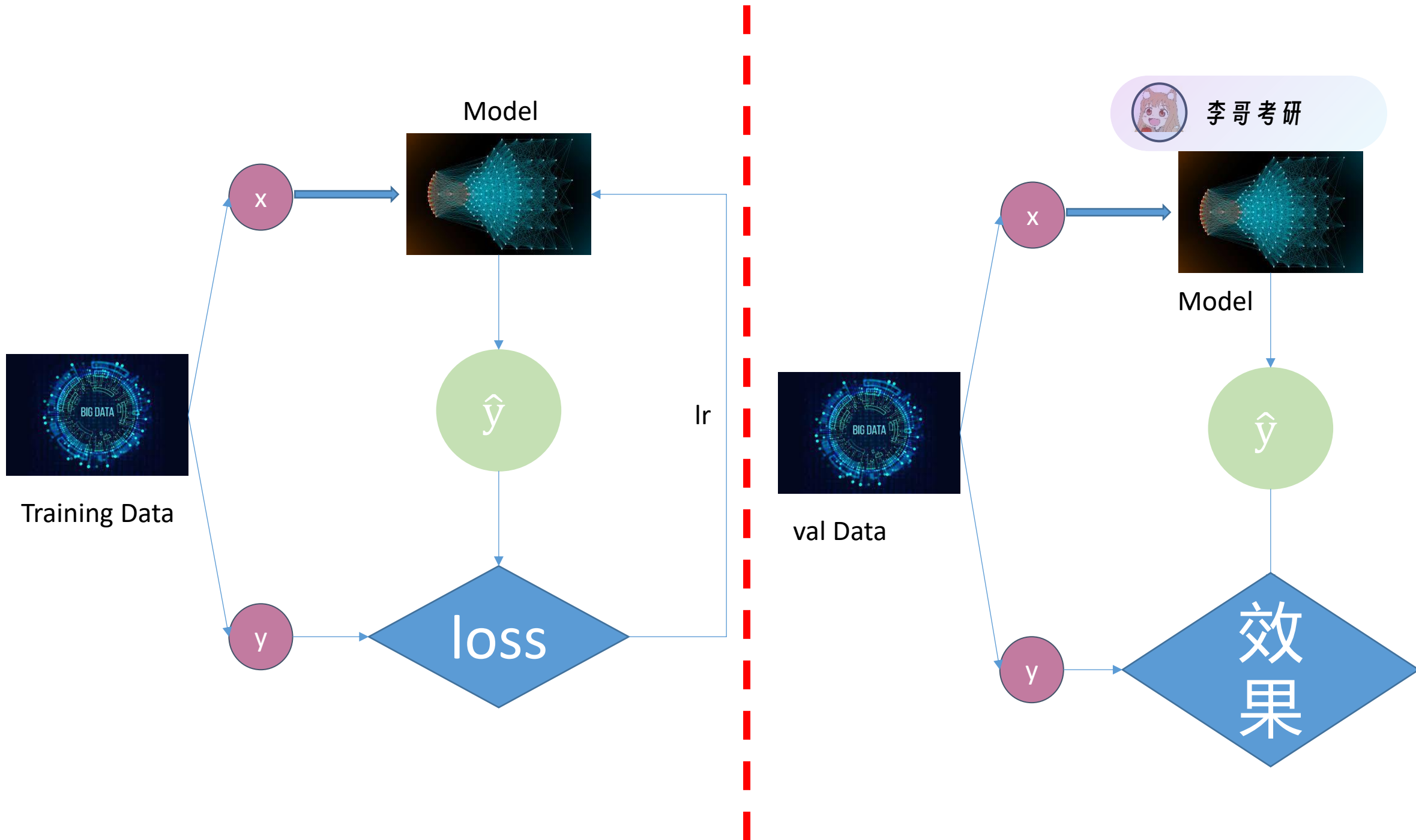
$$\text{update } \theta^2 \leftarrow \theta^1 - \eta \mathbf{g}$$

➤ Compute gradient $\mathbf{g} = \nabla L^3(\theta^2)$

$$\text{update } \theta^3 \leftarrow \theta^2 - \eta \mathbf{g}$$

1 **epoch** = see all the batches once





动量



李哥考研



房价预测

此数据集是King County地区2014年五月至2015年五月的房屋销售信息，适合于训练回归模型。



李哥考研

- https://www.kaggle.com/datasets/harlfoxem/housesalesprediction?select=kc_house_data.csv
- id: 唯一id
- **date**: 售出日期
- price: 售出价格（标签值）
- bedrooms: 卧室数量
- bathrooms: 浴室数量
- sqft_living: 居住面积
- sqft_lot: 停车场面积
- floors: 楼层数
- waterfront: 泳池
- view: 有多少次看房记录
- condition: 房屋状况
- grade: 评级
- sqft_above: 地面上的面积
- sqft_basement: 地下室的面积
- yr_built: **建筑年份**
- yr_renovated: **翻修年份**
- zipcode: **邮政编码**
- lat: 维度
- long: 经度
- sqft_living15: 2015年翻修后的居住面积
- sqft_lot15: 2015年翻修后的停车场面积

思考：
如何处理日期和邮政编码。
Label值太大怎么办
(统一除以一个值)

id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	grade	sqft_above	sqft_basement	yr_built	yr_renovated	zipcode	lat	long	sqft_living15	sqft_lot15
7.129E+09	20141013T00:00:00	221900	3	1	1180	5650	1	0	0	3	7	1180	0	1955	0	98178	47.5112	-122.257	1340	5650

总结



李哥考研

在回归这一章，应该记住的

Linear (4, 1)

Linear (4, 100)

Linear (100, 1)

Loss 与梯度下降



李哥考研

答疑和结束

THANKS



从零设计一个线性回归模型。



李哥考研

理解 代码的每一步是在做什么事情。

回顾昨天的代码。



李哥考研

先找到对应的部分。模型。数据， loss ， 参数， 训练

尝试加深网络层数。

尝试加宽网络。

增加输入维度。

自己写一个数学公式， 设计回归网络， 并计算参数量。