

习题模块

一、线性回归

1.

假如你正在实现一个全连接层，全连接层的输入形状是 7×8 ，输出形状是 7×1 ，其中7是批量大小，则权重参数 w 和偏置参数 b 的形状分别是__和__

- ☐ $1 \times 8, 1 \times 1$
- ☐ $1 \times 8, 7 \times 1$
- ☒ $8 \times 1, 1 \times 1$
- ☐ $8 \times 1, 7 \times 1$

In [13]:

```
1 def linreg(X, w, b):  
2     return torch.mm(X, w) + b
```

参考笔记中文模型： $w \times b$ 的层表： X 为 100×2 ； w 为 2×1 ， b 为 1×1 ；在初始定义 b 时，需要指定size；
这里是广播机制的原理，关键是模型的参数形状不应该和批量大小挂钩；

2.

课程中的损失函数定义为：

```
def squared_loss(y_hat, y):  
    return (y_hat - y.view(y_hat.size())) ** 2 / 2
```

这道题做错了：

主要问题是没有理解清楚广播机制：

将返回结果替换为下面的哪一个会导致导致模型无法训练：（阅读材料：<https://pytorch.org/docs/stable/notes/broadcasting.html>）

- ☒ $(y_hat.view(-1) - y) ** 2 / 2$
- ☒ $(y_hat - y.view(-1)) ** 2 / 2$
- ☐ $(y_hat - y.view(y_hat.shape)) ** 2 / 2$
- ☐ $(y_hat - y.view(-1, 1)) ** 2 / 2$

① 首先先对原损失函数解释下： y_hat 为通过 $Xw+b$ 得到的： 10×1 维的向量；

y 为 $(1,0)$ 的向量；(labels)，而 y 是一个批处理的数据为 $(1,0)$ ≠ $(1,0,1)$

我刚开始写时，觉得取用：

$(y_hat - y) ** 2 / 2$ ；

但是这样的问题在于，通过广播机制， $(y_hat - y)$ 会得到 10×10 的矩阵；而初始的向量是得到 10×1 的；

故使用： $y_hat - y.view(y_hat.size()) \rightarrow 10 \times 1$ ；

② 解释下view函数：

view函数的作用是改变原tensor的张量；

$view(-1)$ 是将tensor平铺，如tensor是 3×4 的，平铺后为 $(12,)$ ；

$view(-1,1)$ 是将tensor转置；但是如果tensor为 $(n,)$ 时，转置后为 $(n,1)$ ；

③ 广播机制 \Rightarrow

一、触发广播机制的条件：

广播机制可以完成不同shape的tensor间的运算，其需要满足两个条件：

- ① 两个tensor的维数大于等于1；
- ② 从后往前遍历两个tensor的维数，对应位置上的数要么相同，要么有一方为1，要么有一方不存在；

如： $torch.empty(5,3,4,1)$ } \rightarrow $\begin{bmatrix} 5 & 3 & 4 & 1 \\ 3 & 1 & 1 & 1 \end{bmatrix}$ 相同，符合广播的条件；
 $torch.empty(3,1,1)$

一、广播机制后的返回值：

- ① 如果两个tensor的维数不相同, 则在维数少的tensor后面补1;
 ② 接着依次比较同一维度的值, 选择最大的值作为该维度的值;

如: $\text{torch.empty}(5, 3, 4, 1)$ } $5, 3, 4, 1$ } $\begin{bmatrix} 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
 $\text{torch.empty}(3, 1, 1)$ } $1, 3, 1, 1$ } $\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$
 不足补1 (5, 3, 4, 1) 输出

④ 首先对A选项: $\underline{y_hat.view(-1)} \rightarrow (10,)$; $\underline{y} \rightarrow (10,)$; 相减后仍为 $(10,)$,
 容易出错, 但是求出的向量, 可以接受;
 B选项: $\underline{y_hat} \rightarrow (10, 1)$, $\underline{y.view(-1)} \rightarrow (10,)$, 两个tensor维数不同, $\begin{bmatrix} 10 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 一个为1, 一个为0, 故符合广播。
 $\begin{bmatrix} 10 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow (10, 10);$

C选项: $\underline{y_hat} \rightarrow (10, 1)$; $\underline{y.view(y_hat.shape)} \rightarrow (10, 1)$; 两个tensor维数相同, 相减为 $(10, 1)$, 符合要求;

D选项: $\underline{y_hat} \rightarrow (10, 1)$; $\underline{y.view(-1, 1)} \rightarrow (10, 1)$; 相减后为 $(10, 1)$, 符合要求;