

第二次作业

写代码生成一个包含 200 条数据的数据集（数据集的具体要求见步骤 1），再按照步骤 2) ~ 步骤 4) 的要求生成指定模型，训练模型、并分别画出训练过程中模型在训练集和验证集上的学习曲线，利用学习曲线判断模型是欠拟合状态、过拟合状态，还是表现正好。

1) 生成一个人工数据集。在训练数据集和测试数据集中，给定样本特征 x ，我们使用如下的三阶多项式函数来生成该样本的标签： $y = 1.2x - 3.4x^2 + 5.6x^3 + 5 + \epsilon$ ；其中噪声项 ϵ 服从均值为 0、标准差为 0.1 的正态分布。训练数据集和测试数据集的样本数都设为 100。

```
m = 200
X = 2 * np.random.rand(m, 1) - 1
y = 5.6*X**3 - 3.4*X**2 + 1.2*X + 5 + np.random.normal(0, 0.1, size=(m,1))
```

2) 先使用与数据生成函数同阶的三阶多项式函数拟合。

①打印出最佳参数的取值，与真实参数值进行对比，并对结果进行分析点评；

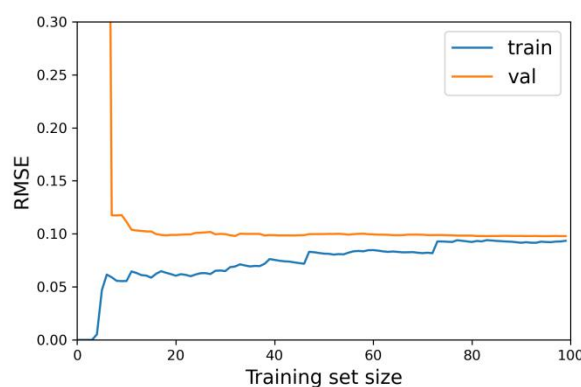
最佳参数取值：[4.99543715] [[1.23762358 -3.3675671 5.53132528]]；

即三阶多项式函数拟合后结果为：

$$y = 1.23762358x - 3.3675671x^2 + 5.53132528x^3 + 4.99543715$$

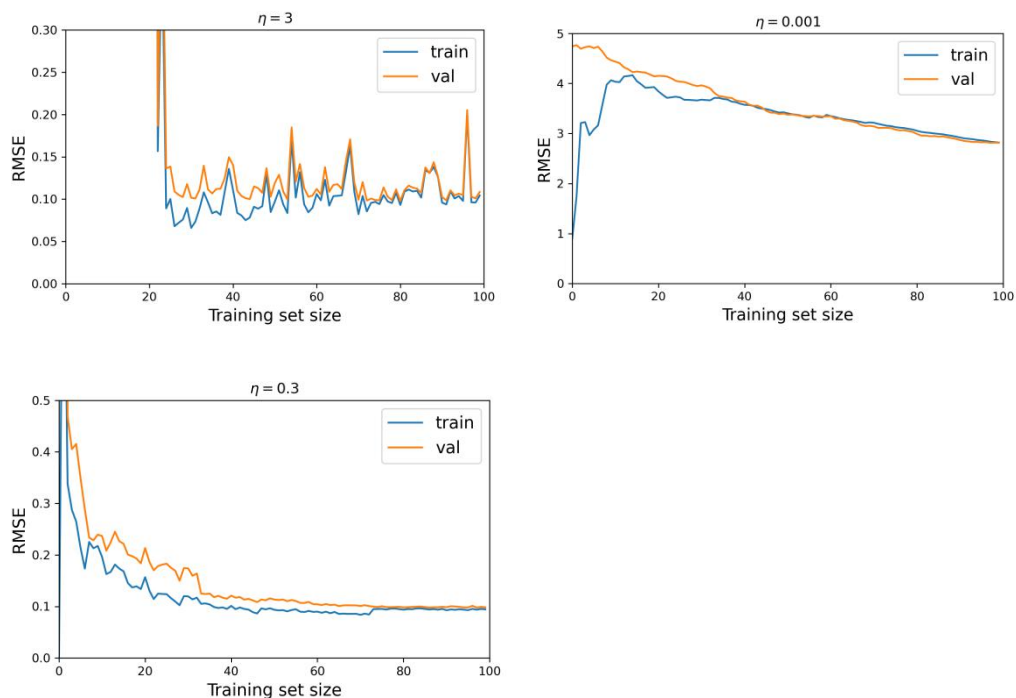
从结果可以看出，最佳参数取值与实际生成数据的函数取值：1.2, -3.4, 5.6, 5 很接近。由于噪声项的存在，拟合后的结果不可能与原结果一致，但是这个模型的训练误差和在测试数据集的误差都较低，训练出的模型参数也接近真实值。

②尝试画出训练过程中模型分别在训练集和验证集上的学习曲线；观察分析这 2 条学习曲线，判断该模型是欠拟合，还是过拟合，还是表现很好？请说明理由。



从模型在训练集和验证集上的学习曲线上可以看出，模型在训练集和验证集上的误差很接近，在训练集上的效果略好于在验证集上的效果，属于正常现象，并且随着训练轮次的增加，误差都下降到一个很小的范围，所以可以模型表现很好。

③尝试使用不同的学习率 η （四种情况：过大，过小，正好），画出不同的学习率 η 下，模型在训练集上的学习曲线；对比观察这三条学习曲线的走势，分析学习率对训练过程的影响。

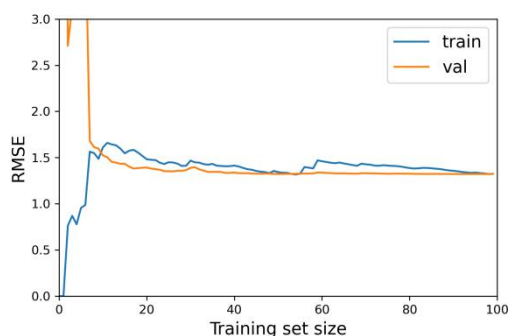


如图，分别画出了学习率为分别为 3、0.001、0.3 的学习曲线。

1. 学习率为 3 时，此时学习率过大，学习的速度很快，梯度很早下降，但是算法过于发散，学习曲线总体处于震荡的趋势。
2. 学习率为 0.001 时，此时学习率过小，学习速度很慢，梯度下降不明显，收敛耗时过长，训练完 100 条数据后误差还不足以达到一个令人满意的状态。
3. 学习率为 0.3 时，此时学习率正好，模型在训练集和验证集的表现比较接近，能在较短时间内收敛，收敛后的误差较小。

3) 再试试直接用线性函数拟合。

尝试画出训练过程中模型分别在训练集和验证集上的学习曲线；观察分析这 2 条学习曲线，判断该模型是欠拟合，还是过拟合，还是表现很好？请说明理由，并指出改进方案。



用线性函数拟合时，该模型欠拟合。因为该模型的训练误差在迭代早期下降后便很难继续降低，RMSE 一直维持在一个较高的水平，训练误差很高。

改进方案：可以选用更复杂的模型。

4) 使用三阶多项式函数模型来拟合, 请注意, 这次, 只使用2个样本来训练模型。

①打印出最佳参数的取值, 并与真实参数值进行对比, 并对结果进行分析点评。

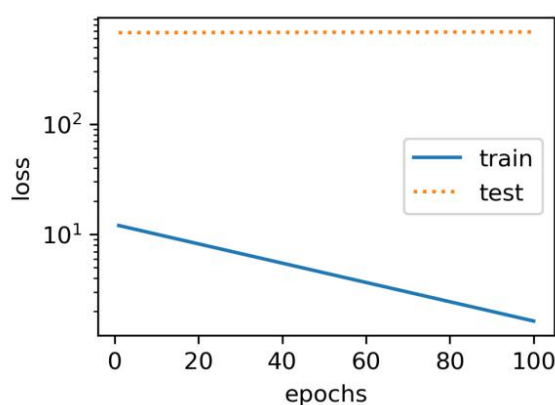
最佳参数取值: -0.2118931, 0.03864809, -0.03436531, 3.1383252;

即三阶多项式函数拟合后结果为:

$$y = -0.2118931x + 0.03864809x^2 - 0.03436531x^3 + 3.1383252$$

从结果可以看出, 最佳参数取值与实际生成数据的函数取值: 1.2, -3.4, 5.6, 5 差距很大。说明该模型在只使用 2 个样本来训练时效果特别差, 很有可能出现了过拟合现象。

②尝试画出训练过程中模型分别在训练集和验证集上的学习曲线; 观察分析这 2 条学习曲线, 判断该模型是欠拟合, 还是过拟合, 还是表现很好? 请说明理由, 并指出改进方案。



(使用的是《动手学深度学习》的代码绘制的图)

只使用 2 个样本来训练时, 该模型过拟合。

只使用两个样本来训练模型。显然训练样本过少了, 甚至少于模型参数的数量。这使模型显得过于复杂, 以至于容易被训练数据中的噪声影响。在迭代过程中, 尽管训练误差较低, 但是测试数据集上的误差却很高。

改进方案: 添加更多的样本来训练模型。

5) 使用40阶多项式函数模型来拟合。

①打印出最佳参数的取值, 并与真实参数值进行对比;

最佳参数取值为:

常数项: [5.01677319]

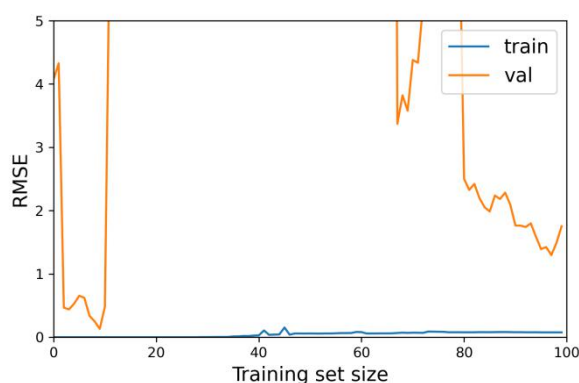
从一次项系数到 40 次项系数:

```
[[ 2.32492409e-01 -2.30394132e+01  2.60829658e+02  3.66839742e+03
 -2.04871895e+04 -2.21822869e+05  7.79381473e+05  6.78701386e+06
 -1.72001412e+07 -1.25430347e+08  2.44916780e+08  1.53995209e+09
 -2.40039234e+09 -1.33186667e+10  1.69113837e+10  8.43514615e+10
 -8.82902874e+10 -4.01683980e+11  3.48902958e+11  1.46433076e+12]
```

-1.05847134e+12 -4.13390937e+12 2.48505953e+12 9.09271379e+12
 -4.52622335e+12 -1.55975163e+13 6.37385340e+12 2.07731807e+13
 -6.86901336e+12 -2.12449662e+13 5.55673304e+12 1.63560388e+13
 -3.26401071e+12 -9.16579572e+12 1.31379993e+12 3.52734450e+12
 -3.23975280e+11 -8.33264124e+11 3.68959941e+10 9.10731986e+10]]

实际生成数据的函数 $y = 1.2x - 3.4x^2 + 5.6x^3 + 5 + \varepsilon$ 为 3 次函数，没有可比性。

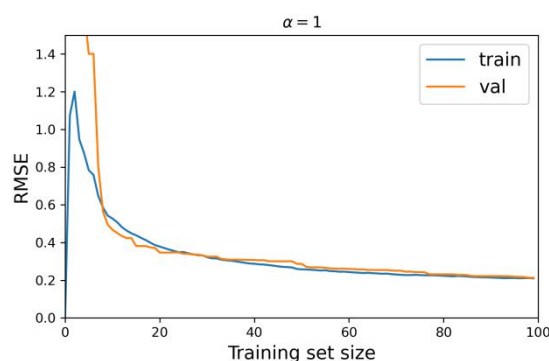
②尝试画出训练过程中模型分别在训练集和验证集上的学习曲线；观察分析这 2 条学习曲线，判断该模型是欠拟合，还是过拟合，还是表现很好？请说明理由，并指出改进方案。



画出训练过程中模型分别在训练集和验证集上的学习曲线，可以看出该模型过拟合。因为这两条曲线之间有较大差距，这意味着该模型在训练集上的表现比验证集上要好很多，这是过拟合的标志。而且从学习曲线可以看出，在验证集上，RMSE 曲线波动很大，可能是使用的模型太过复杂，容易造成过拟合。

改进方案：提供更多的训练数据或者对模型正则化。

③尝试利用 L2 正则化技术，让该模型的表现好起来。画出训练过程中模型分别在训练集和验证集上的学习曲线；观察分析这 2 条学习曲线，判断该模型是欠拟合，还是过拟合，还是表现很好？并说明理由。



利用 L2 正则化技术后，该模型表现得很好。因为随着样本量的增大，模型在训练集和验证集上最终都收敛了，而且对应的 RMSE 都很小，两条曲线之间的差距也很小。

④打印出③中模型的最佳参数，并与①进行对比，尝试总结下 L2 正则化技术所带来的效果，并分析为何 L2 正则化技术可以解决过拟合问题。

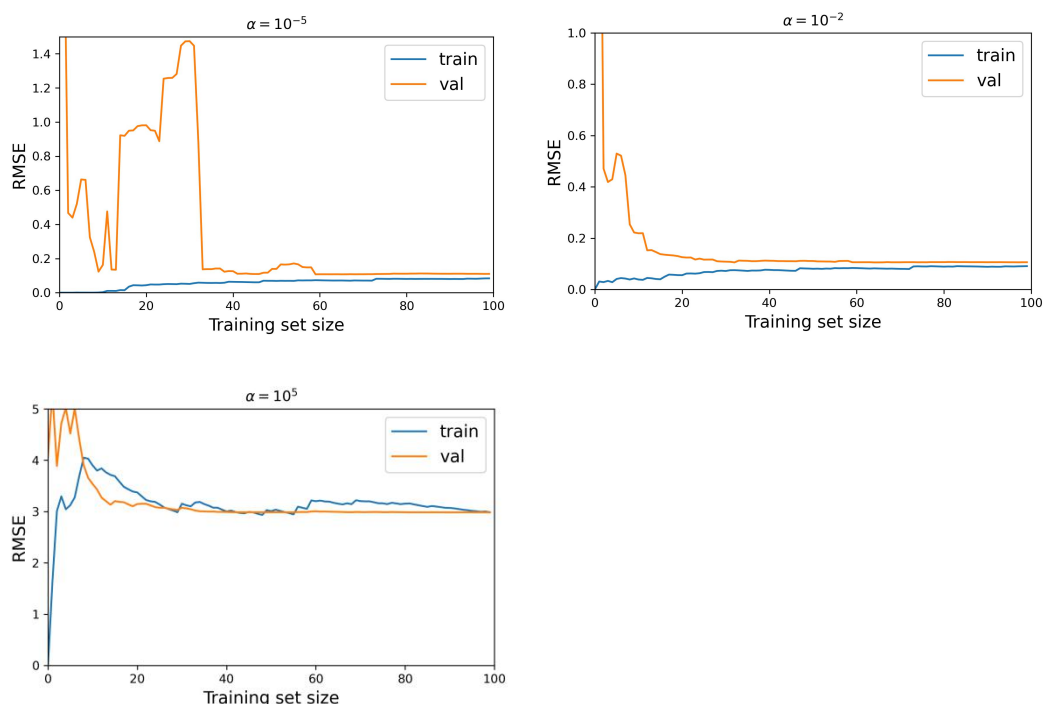
4.792648181511375

```
[ 2.19019975 -1.99792592  2.13146801 -0.99009263  1.3790219  -0.45572006
  0.84471717 -0.18625306  0.50412577 -0.04694692  0.28893372  0.02561486
  0.15216527  0.0622642   0.06477961  0.07893559  0.00896476  0.08430181
 -0.02634555  0.08326694 -0.04814978  0.07867878 -0.06095565  0.07221998
 -0.0677178   0.06489397 -0.07040445  0.05729979 -0.07034595  0.04979179
 -0.06845309  0.04257455 -0.06535623  0.03576026 -0.0614954   0.02940401
 -0.05717966  0.02352561 -0.05262678  0.01812335]
```

与①中的参数进行对比，可以看出使用 L2 正则化之后模型的各项参数都变得比较小。

对于线性模型来说，正则化通常通过约束模型的权重来实现。正则化通过为模型损失函数添加惩罚项使学出的模型参数值较小，以此来应对过拟合。L2 范数正则化在模型原损失函数基础上添加 L2 范数惩罚项，从而得到训练所需要最小化的函数，通过惩罚绝对值较大的模型参数为需要学习的模型增加了限制，这可能对过拟合有效。

⑤ 在③中使用不同的正则化强度，打印出不同正则化强度下的最佳参数，并对结果进行分析点评。



对应的参数分别为：

5.019169830216255

```
[ 0.93449821 -3.63673426  10.24972196 -0.85301067 -17.83774384
  8.44127224  17.48497098 -10.94912959  8.70512859  -5.16181598
 ...
 1.9060812  -0.32385947 -1.67078619  4.22147988 -6.87952094]
```

```

5.019169830216255
[ 0.93449821 -3.63673426 10.24972196 -0.85301067 -17.83774384
 8.44127224 17.48497098 -10.94912959 8.70512859 -5.16181598
...
1.9060812 -0.32385947 -1.67078619 4.22147988 -6.87952094]

4.987924548021155
[ 1.43707121e+00 -3.26740161e+00 4.43482296e+00 -1.20964883e-01
 1.20218427e+00 -4.41080845e-02 2.15152274e-01 -1.76424960e-01
...
1.45674019e-01 -2.64045562e-01 1.73319158e-01 -3.24990982e-01]

```

超参数 α 控制的是对模型进行正则化的强度，可以看出 α 越大，模型的参数就越小，如果 α 非常大，那么所有的权重都将非常接近于 0。

从画出的学习曲线中也可以看出，当 $\alpha=10^5$ 时，此时由于正则化的程度太高，大大降低了模型的复杂度，即使是原来 40 阶的多项式模型此时也出现了欠拟合的状态。当 α 在 0.01 时，此时模型的表现很好，有效降低了原本模型的复杂度，也没有出现欠拟合。

提示：可参照《动手学深度学习》的“3.11.4 多项式函数拟合实验”