

L10T4

☐ 改变算法中的各类超参数、样本数量、样本分布等，对于梯度下降法还要改变不同的学习率以及不同的batch size和不同epoch次数，讨论实验结果。

分布一

样本分布：

- +1类：均值【3， 0】 协方差：单位矩阵 数量：200
- -1类： 均值【0， 3】 协方差：单位矩阵 数量：200

数据记录

- 使用随机种子控制随机性
- 记广义逆方法为NE(Normal-Equation)，梯度下降法为GD(Gradient-Descent)

	epoch_num=1000				epoch_num=10			
	lr=0.25	lr=0.2	lr=0.1	lr=0.005	lr=0.25	lr=0.25	lr=0.1	lr=0.005
test_acc_NE	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
test_acc_GD	0.0(已发 散)	0.5	1.0	1.0	0.5	0.5	1.0	0.7

分布二

样本分布

- +1类：均值 [4, 2] 协方差： $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ 数量：75
- -1类： 均值 [0, 4] 协方差： $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 数量：75

数据记录

- 使用随机种子控制随机性
- 记广义逆方法为NE(Normal-Equation)，梯度下降法为GD(Gradient-Descent)

	epoch_num=1000				epoch_num=10			
	lr=0.1	lr=0.05	lr=0.01	lr=0.005	lr=0.1	lr=0.05	lr=0.01	lr=0.005
test_acc_NE	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
test_acc_GD	0.0(已发散)	0.93	0.93	0.93	0.43	0.93	0.13	0.13
plot								

结论分析

1. 由于线性回归优化问题存在解析解，可用广义逆法直接求出理论最优解 w ，此时就是数据决定学习上限。如分布一中，由于样本分布可分性良好，因此理论上的模型在测试集上的准确度为1，由于分布二样本分布可分性较好但不算完美（如下图所示），因此理论上的模型在测试集上的准确度到不到1，广义逆法拟合的准确率只有0.93。
2. 梯度下降法是各种算法的通用优化算法，在一次次迭代中不断调整参数，因此在合适的学习率与足够多的迭代次数下，梯度下降法可以逼近优化的理论上限（也就是达到广义逆法的效果）。但如果超参数设置不合适，会出现其他问题：
 - a. 在本实验中，较为理想的学习率为0.05~0.1，超过0.1，易导致算法不收敛甚至发散到无穷远处；
 - b. 如果学习率较小（如0.005），则模型学习速率较慢，模型效果依赖迭代轮次，只要epoch足够大，模型总能学习到最优解，原因就是线性回归中的损失函数为均方误差，是凸函数；而如果epoch不够大，则模型还没收敛就停止更新了，此种情况下模型效果也较差。

Scatter Plot of Data Points

