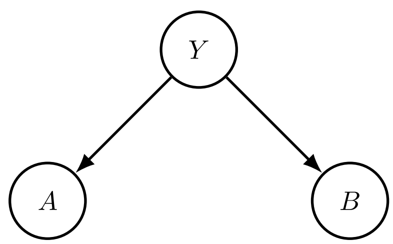
作业 5，人工智能导论课（2023 春季学期）

朴素贝叶斯, 感知机，罗吉斯特回归优化

朴素贝叶斯

我们将用朴素贝叶斯模型，基于两个特征变量 A 和 B，来判断标签变量 Y。所有这三个变量都是二元变量，值域都是集合 $\{0,1\}$。以下给出了 10 个训练样本，这些样本点将用来估计模型参数。



A	B	Y
1	1	1
1	0	1
1	0	0
1	1	0
0	1	0
1	1	1
0	1	1
1	0	0
1	1	0
1	1	0

1. 请计算该模型的概率分布表的最大似然法估计值是多少?

A	Y	$P(A\|Y)$
0	0	
1	0	
0	1	
1	1	

B	Y	$P(B\|Y)$
0	0	
1	0	
0	1	
1	1	

Y	$P(Y)$
0	
1	

2. 给定一个新的测试样本 $(A=1, B=1)$，请计算该模型对这个样本的预测标签是什么?

3. 请应用 Laplace 平滑方法重新计算概率分布 $P(A|Y)$ 的值，假定 Laplace 平滑参数 $k=2$。

A	Y	P(A|Y)
0	0	
1	0	
0	1	
1	1	

感知机

假设给定以下的 5 个编号的数据向量点，每个数据向量由 A 和 B 两个维度构成，标签列给出的是对应数据向量的分类，假设是二分类的情况。请回答以下各个小题。

数据索引	A	B	分类标签
1	1	1	-
2	3	2	+
3	2	4	+
4	3	4	+
5	2	3	-

1）检查上面这些不同类别的数据点是否可以被线性分割辨别出来?

2）训练一个感知机分类器。假设初始权值向量是 [-1,0,0]。依次使用上面给出的数据训练一轮该感知机模型，得到的权值向量是什么？假设我们对所有的输入数据都增加一维的常量值，即 $f_0 = 1$，$f_1 = $ A 的值 $f_2 = $ B 的值.

3）在 2）中得到的感知机是否可以正确分类以上数据?

罗吉斯特回归

我们要训练一个数据分类判别器，假设有 N 个训练样本数据，每个样本数据包括一个特征向量 $\mathbf{x} = \{x_1, ..., x_k\}$, 和一个标签 $y \in \{0, 1\}$，使用的分类器的模型是罗吉斯特回归 (logistic regression)，它所给出的预测是基于以下的概率函数：

$$P(Y=1|X) = h(\mathbf{x}) = s(\sum_i w_i x_i) = \frac{1}{1+\exp(-(\sum_i w_i x_i))}$$

用 $s(\gamma)$ 表示罗吉斯特函数，并且 $\exp x = e^x$，即：

$$s(\gamma) = \frac{1}{1+\exp(-\gamma)}$$

这个模型所要求解的权值参数是：$\mathbf{w} = \{w_1, ..., w_k\}$.

优化方法将使用随机梯度下降法 (stochastic gradient descent) 求解最优的权值 w_j。给定那 N 个训练样本数据，想要最小化以下的损失目标函数 (loss function)：

$$L = -[y \ln h(\mathbf{x}) + (1-y)\ln(1-h(\mathbf{x}))]$$

1) 请推导出损失函数相对于 w_i 的偏导函数，即 $\frac{dL}{dw_i}$。提示：$s'(\gamma) = s(\gamma)(1-s(\gamma))$。

2) 借助上面推导出的梯度结果，请写出使用随机梯度下降法更新 w_i 的表达公式。假设步长参数用 η 来表示。