第二周作业

第三讲作业

姓名：乔翱

学号：202222080625

1、用线性回归处理非线性问题有什么办法，试举其中三个具体例子。

通过采用基函数，对原始数据变量进行一些预处理或特征提取，使得线性模型可以是输入变量的非线性函数。常见的基函数有多项式基函数、高斯基函数和sigmoid基函数。

将原始特征进行转换，转换为新的特征，建立线性关系。

使用核技巧，将低维特征映射到高维特征空间中，将原始特征映射到高维空间。

2、机器学习按任务可以分成哪三类？机器学习的三要素分别是什么？用最小二乘法（Least Squares Regression）拟合一条直线，试给出其三个要素，并拟合下面三个数据点：(1, 1.1)、 (2, 2.1)、 (3, 2.8)。

机器学习按任务可以分成监督学习、无监督学习和强化学习。

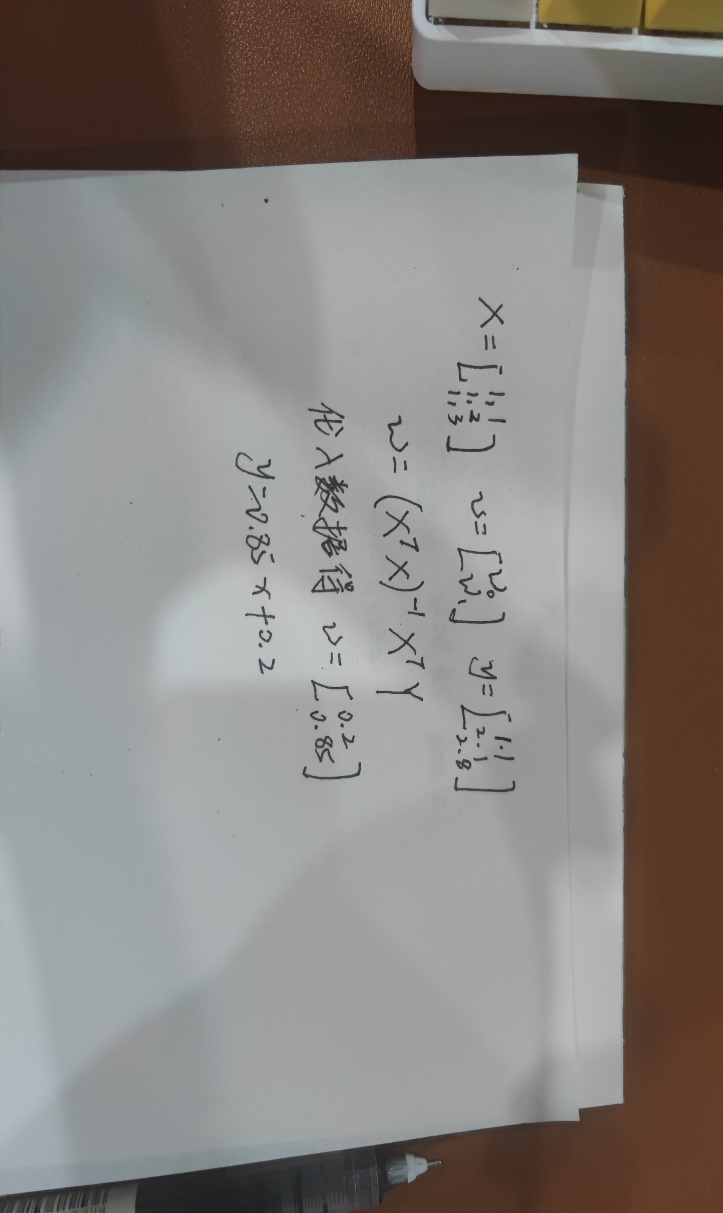
机器学习的三要素是模型、学习准则和优化方法。

最小二乘法拟合一条直线：

模型：线性模型

学习准则：损失函数为

优化方法：批量梯度下降，随机梯度下降等，最小二乘法可以直接求解



3、什么是最小平方回归（Least Squares Regression），写出它的损失函数，试用从概率观点给出解释。

最小平方回归是通过最小化预测值和真实值之间的平方误差，来拟合出一个好的线性模型。

损失函数

从概率角度看，最小平法回归可以看作是假设y与x之间存在一种线性关系，并且y是带有高斯噪声的。最小平方回归的目标是最大化条件概率p（y|x），在给定x的情况下，y的条件概率服从高斯分布，使得观测到的数据点的似然最大化。

4、什么是岭回归（Ridge Regression），写出它的损失函数，试用从概率观点给出解释，并说明它的优点。

岭回归通过对损失函数增加一个正则化项来缓解过拟合问题，正则化项是模型参数的平方和，并且引入一个正则化参数α控制正则化强度。损失函数为J(θ)=1/2(Xθ−Y)T(Xθ−Y)+1/2α||θ||22

从概率的角度，岭回归可以看作是在最小化损失函数的同时，对模型参数w进行约束，可以看作是引入了一个先验概率分布，而正则化项对应于先验概率的负对数概率。

特点：缓解过拟合，提高模型泛化能力，对多重共线性问题处理，减少特征之间的相互影响，有效防止过拟合。

5、什么是LASSO回归，写出它的损失函数，试从概率观点给出解释，并说明它的特点。

LASSO回归在最小化平方误差的同时，对模型参数进行约束，以实现特征选择和模型简化的目的，其损失函数为J(θ)=1/2n(Xθ−Y)T(Xθ−Y)+α||θ||1

从概率论的角度看，LASSO回归可以看作是对模型参数引入一个拉普拉斯先验分布，这个分布在中心点处有一个尖峰，使得大部分参数都趋于0，只有极少数的参数取非零。因此,LASSO回归可以自动进行特征选择，将一些不重要的特征对应系数收缩到0，从而减少了模型的复杂度，提高模型的泛化能力。简要概况，LASSO回归会导致模型参数稀疏，进行变量选择，降低过拟合。

特点：进行特征选择，减少特征对模型的影响，提高了模型泛化能力，将一些参数收缩到0，从而进行变量选择，获得更加紧凑的模型，同时降低过拟合的风险。

第四讲作业

1. 简述解决分类问题的三种方法，并各举1种具体的方法。

判别函数：直接找到一个函数f（x），把每个输入x直接映射为类别标签。感知器

概率生成模型：先对类条件密度和先验类概率分布建模然后再使用贝叶斯定理计算后验类概率分布，最后，使用决策论来确定每个输入的类别。等价，直接对联合概率分布建模，再归一化得到后验概率。朴素贝叶斯

概率判别式模型：直接对后验概率建模的方法，再使用决策论来确定每个新的输入x的类别。逻辑回归、BP神经网络。

1. 请给出Logistic回归分类模型的三要素。

假设函数：逻辑回归使用sigmoid函数作为其假设函数，之后将线性函数的结果映射到[0,1]的概率范围内，来表示样本属于某一类的概率。

损失函数：逻辑回归使用对数损失函数作为损失函数，用来度量模型预测结果与实际结果之间的差异。

参数优化算法：逻辑回归使用梯度下降算法来最小化损失函数，梯度下降算法通过反复迭代更新模型参数，最终使得损失函数逐渐收敛到最小值。

1. 朴素贝叶斯。

考虑一个二分类问题，类别标签，每个训练样本有两个属性，假设*X*1和*X*2在个给定*Y*情况下条件独立，类别先验，已知条件概率如下：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *P*(*X*1|*Y*) | *X*1 = 0 | *X*1 = 1 |  | *P*(*X*2|*Y*) | *X*2 = 0 | *X*2 = 1 |
| *Y* = 0 | 0.2 | 0.8 |  | *Y* = 0 | 0.5 | 0.5 |
| *Y* = 1 | 0.7 | 0.3 |  | *Y* = 1 | 0.9 | 0.1 |

利用朴素贝叶斯方法预测*X*1和*X*2四种可能取值对应的标签，并填入下表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *X*1 | *X*2 | *P*(*X*1, *X*2*|Y* = 0) | *P*(*X*1, *X*2*|Y* = 1) | (*X*1, *X*2) |
| 0 | 0 | 0.1 | 0.63 | 0.365 |
| 0 | 1 | 0.1 | 0.07 | 0.085 |
| 1 | 0 | 0.4 | 0.27 | 0.335 |
| 1 | 1 | 0.4 | 0.03 | 0.215 |