

1.模型的评价指标？这些评价指标是如何计算、获取？

(1)

分类模型的评价指标：

TP(True Positive)真阳性：预测为正，实际也为正；

FP(False Positive)假阳性：预测为正，实际为负；

FN(False Negative)假阴性：预测为负，实际为正；

TN(True Negative)真阴性：预测为负，实际也为负。

精确率 P：

$$P = \frac{TP}{TP + FP} \quad (1)$$

召回率 R：

$$R = \frac{TP}{TP + FN} \quad (2)$$

(2)

回归问题的评价指标：

①平均绝对误差：

$$MAE = \frac{1}{m} \sum_{i=1}^m (|y_i - f(x_i)|) \quad (3)$$

②均方误差：

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2 \quad (4)$$

③均方根误差：

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2} \quad (5)$$

式(3)-(5)中的 y_i 为实际值， $f(x_i)$ 为模型预测值。

可以根据以上误差的大小对模型的参数进行调整，从而选择更好的模型。

2.分类问题、标注问题、回归问题是怎么定义的？解决分类问题、标注问题、回归问题常用的统计学习方法有哪些？

(1)

分类问题的输出是有限个离散值，每个类别的置信度。

常用的统计学习方法：

近邻法、感知机、朴素贝叶斯法、决策树、logistic 回归、SVM、adaBoost 等。

(2)

标注问题的输入是一个观测序列，输出的是一个标记序列或状态序列。例如在自然语言处理中给定一个由单词组成的句子，对这个句子中的每一个单词进行词性标注就属于标注问题。

常用的统计学习方法：

隐性马尔可夫模型、条件随机场。

(3)

回归问题表示从输入变量到输出变量之间映射的函数，预测输入变量与输出变量之间的关系。

常用的统计学习方法：

神经网络

3.请描述一个完整的统计学习方法步骤。

- 1.将数据集进行划分 (train、val、test)；
- 2.明确要学习的是哪类模型；
- 3.确定模型选项的准则，即确定损失函数；
- 4.选择最优模型的算法，即按照什么计算方法求解最优模型；
- 5.选择最优模型；
- 6.利用最优模型对新数据进行预测。