

## 作业 2

1. 答案略。注意读题，问的是两个问题分别列举两种代表性算法 (respectively)，有的同学只列了一种。
2. 答案为 49/27。三次交叉验证的结果取平均即可。
3. 第一问为  $(y, x+2y)$ ，第二问为  $(1/9, 2/9, -2/9)$ 。本题多为计算错误。
4. 最终决策树



(After 8 处用 wind 作为划分属性亦可)

错得最多的地方在于 Wind 属性下漏掉 no wind 分支。对于这种情况，群内有过讨论，可阅读决策树构建的源码寻找答案。

5. 预测为-1。本题不需要使用拉普拉斯变换（没有遇到属性被抹去的情况）。

作业 3

2.  $55 \times 55 \times 96$

3. 本课程不需要大家考虑卷积核翻转的问题。这里唯一要注意的就是计算准确度的问题。

(1)

(a)

15	16
6	15

(b)

7	12	10	2
4	15	16	10
10	6	15	6
8	10	4	3

(2)

Max

8	4
7	5

Average

4.5	2.5
4.25	3

## 模拟题

1. In a least-squares linear regression problem, adding an L2 regularization penalty cannot decrease the L2 error of the solution  $w$  on the training data.

True. 关键词在于 decrease 和 training data。L2 error 虽然课上没有明确讲到，但类比 L2 范数，可以猜到就是平方误差。

2. Which of the following statements is/are correct?
- A. Cross validation will guarantee that our model does not overfit.
  - B. Logistic Regression can be used for classification.
  - C. Increasing the depth of a decision tree cannot increase its training error.
  - D. In the discriminative approach to solving classification problems, we model the conditional probability of the labels given the observations.

BCD。漏选 C 的同学最多，树越深，拟合能力越强，理论上 training error 减小。

3. The training error of 1-NN classifier is \_\_\_\_\_.

0. 小测中也出现过，这就是自己预测自己，误差自然为 0。

4. Say you have an input image whose shape is  $128 \times 128 \times 3$ . You are deciding on the hyperparameters for a Convolutional Neural Network; in particular, you are in the process of determining the settings for the first Convolutional layer. Compute the output activation volume dimensions and number of parameters of each of the possible settings of the first Convolutional layer, given the input has the shape described above. You can write the activation shapes in the format (H, W, C) where H, W, C are the height, width, and channel dimensions, respectively.

(1) The first Conv layer has a stride of 1, a filter size of 3, input padding of 0, and 64 filters.

(2) The first Convolutional layer has a stride of 1, a filter size of 5, input padding of 2, and 16 filters.

(3) The first Convolutional layer has a stride of 2, a filter size of 2, input padding of 0, and 32 filters.

- (1) (126, 126, 64), 1792   (2) (128, 128, 16), 1216   (3) (64, 64, 32), 416

这里没有显式说不要 bias，应该考虑 bias；Padding 0 的意思是不进行 Padding。

## 小测

小测的完整答案不会提供，这里我挑出错得最多的地方给大家一些提示。

1. 训练集 accuracy 低，增加数据还是减少数据更可能让**训练集**准确率提升？

从过拟合的角度考虑，数据少更容易过拟合，训练集准确率更有可能提升。当然，此时测试集准确率更有可能下降。同样的道理，另一问中的选项：

“Retraining (a linear regression model) while discarding some training samples will likely reduce the gap between the train MSE and the test MSE.”

是错误的说法。

**过拟合相关的内容非常重要，请大家好好理解。**

2. 选项：交叉验证能减小训练误差。

这是错误的说法，交叉验证可以用来选出过拟合程度较小的模型，但其不能提高模型自身的性能。

3. 对于 KNN 来说，随着维度的增加，为了维持接近的 performance，数据量需要指数型增长。可参考 KNN 课件中有关维数灾难的图示。

4. 注意“线性回归”和“线性可分”不是一回事。因此选项：

“A linear regressor works very well if the data is linearly separable.”

是错误的说法。

5. 选项：KNN 不能用于回归。

这是错误的说法。除了课上讲的分类以外，KNN 也可以用来做回归，比如通过计算数据点最临近数据点平均值而获得预测值。这种问题需要大家自行拓展思考得出结论。同样，决策树亦可用于回归任务。

6. Consider the following joint distribution on  $X$  and  $Y$ , where  $X \in \{-1, 0, 1\}$  and  $Y \in \{0, 1\}$ :  $p(X = -1, Y = 0) = 0.05$ ,  $p(X = -1, Y = 1) = 0.05$ ,  $p(X = 0, Y = 0) = 0.1$ ,  $p(X = 0, Y = 1) = 0.1$ ,  $p(X = 1, Y = 0) = 0.3$ ,  $p(X = 1, Y = 1) = 0.4$ . You learn that  $X \geq 0$ . What is the largest probability of being correct you can achieve when predicting  $Y$  in this case?

这一题有三分之一的同学做错，建议重新计算一遍看看。答案约为 0.556，计算出分数后自行比对是否正确。

## 复习建议与提示

- 复习过程中，要确保作业题、课件上的例题能够**熟练掌握**，尤其是一些计算题，课件中有步骤，每一步都要知道如何计算。
- 计算的准确性问题应该引起大家重视，不要在这种问题上失分。
- 不必把问题复杂化，若有特殊情况一定会有相对明显的提示。
- 务必读清楚题目，比如要求保留几位小数，`correct/incorrect/wrong`/等。
- 特别注意提交的要求，比如作业命名规则等。如果大家不按命名规则来，提交文件就没法统一处理。感谢大家的配合。