

卷35：机器学习

1. 监督式学习模型通常分为生成类模型（generative models）和判别类模型（discriminative models），下面哪个模型属于生成类模型？

- ✔ A. 朴素贝叶斯(Naive Bayes)
- ✘ B. 逻辑回归(logistic regression)
- ✘ C. 决策树(decision tree)

解答：如果模型是直接求解 $P(Y|X)$ ，则是判别模型；如果模型是通过求解 $P(Y, X)$ 来得到 $P(Y|X)$ ，则是生成模型。选A。

2. 决策树的某个节点上有6个阳性样本(positive samples)以及4个阴性样本(negative examples)，那么这个节点的基尼不纯(Gini impurity, Gini Index)是多少？

- ✘ A. 0.4
- ✔ B. 0.48
- ✘ C. 0.60

解答：

$$\text{Gini} = 1 - \sum_{i=1}^n p_i^2 = 1 - (6/10)^2 - (4/10)^2 = 1 - 0.36 - 0.16 = 0.48$$

拓展阅读：[Gini指数](#)、[Gini系数](#)、[Gini不纯是一回事吗？](#)

3. 关于Adaboost分类模型，下面说法错误的是

- ✘ A. adaboost训练多轮决策树，但是最终结果中每个决策树的权重不同
- ✔ B. adaboost用第 i 轮决策树的训练残差来训练第 $i + 1$ 棵树
- ✘ C. 在最终模型中，精度越高的决策树权重越大

解答：B是错误的。B是GradientBoost的特征。

4. LDA(Linear discriminant analysis)和PCA一样，都可以用来做非监督的数据降维。

- ✘ A. 对
- ✔ B. 不对

解答：LDA是有监督学习算法，可以用来降维。

5. 对于一个二元分类问题，我们采用非线性核的SVM（比如RBF SVM），如何设置模型超参C，可以保证模型线性可分？

- ✔ A. $C = \infty$
- ✘ B. $C = 1$
- ✘ C. $C = 0$

解答：C是惩罚系数，C越大，模型对错误点的容忍度越小。所以选A。