MA234 Homework 1

Duolei WANG, SID:12012727

wangdl2020@mail.sustech.edu.cn

1. Finished.

2. (a) Considering

$$H(\text{Appealing}) = -p(\text{Yes})\log_2 p(\text{Yes}) - p(\text{No})\log_2 p(\text{No}) = 1$$

   (b) Considering

$$\begin{aligned} H(\text{Taste}) &= -p(\text{Salty})\log_2 p(\text{Salty}) - p(\text{Sweet})\log_2 p(\text{Sweet}) - p(\text{Sour})\log_2 p(\text{Sweet}) \\ &= -0.3 \cdot \log_2(0.3) - 0.4 \cdot \log_2(0.4) - \log_3(0.3) \\ &= 2.15 \end{aligned}$$

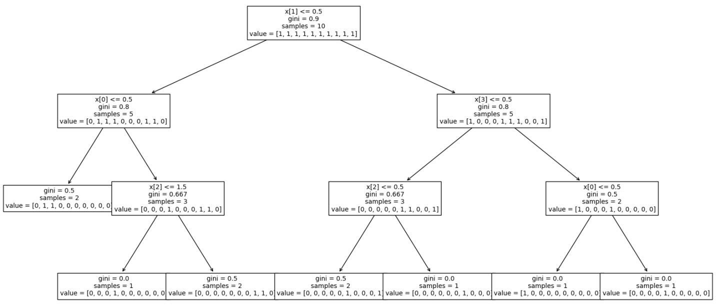
   and

$$p(\text{Salty})H(\text{Appealing}|\text{Salty}) + p(\text{Sweet})H(\text{Appealing}|\text{Sweet}) + p(\text{Sour})H(\text{Size}|\text{Sour})$$

   one can calculate it is 0.4.

   Thus $Gain = 1.75$.

   (c) Considering using the sklearn.tree, with criterion entropy to build a tree, we can get



3. (a) Considering the MLE with μ

$$\mu = \arg\max_{\mu} L(\theta)$$

   Considering σ as a constant, one can derive

$$\mu = \arg\min \sum_{i=1}^{n}(\mu - x_i) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

   (b) Considering optimize the ML with σ, one can derive

$$\sigma = \arg\max_{\sigma} L(\theta)$$

   Derivate the formula with variable σ, use the notation s to representate $\sum_{i=1}^{n}(x_i - \mu)^2$.

   One can get

$$\frac{s - n\sigma^2}{\sigma^{n+3}} \implies \sigma^2 = \frac{s}{n} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

   Use the the MLE $\hat{\mu}$ to replace μ, one can get the hypothesis.

(c) Considering

$$
\begin{aligned}
E(\hat{\mu}) &= \frac{1}{n}E(\sum_{i=1}^{n} x_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} E(x_i) \\
&= \frac{1}{n} \cdot n\mu = \mu
\end{aligned}
$$

Considering

$$
E(\hat{\sigma}^2) = \frac{1}{n-1}E([\sum_{i=1}^{n}(x_i - \bar{x})]^2)
$$

And $(\frac{(x_i-\mu)}{\sigma})^2 \sim \mathcal{X}^2(n)$, thus

$$
E = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2
$$

4. (a) Considering writing the likelihood fuction by $L(\theta) = P_X P_Y$, thus the MLE of p_k is equivalent to the MLE of P_Y. And as y is a discrete variable, one can easily get the rsult as

$$
\hat{p}_k = \frac{Count(y=k)}{n} = \frac{\sum_{i=1}^{n} I(y_i = k)}{n}
$$

(b) Use the MLE of p_k to estimate the p_{sk}, one can note that

$$
\hat{p}_{sk} = \frac{Count(x=s|y=k)}{Count(y=k)} = \frac{\sum_{i=1}^{n} O(x_i = s, y_i = k)}{\sum_{i=1}^{n} I(y_i = k)}
$$

5. Considering the probability of $1-NN$ got a wrong answer, noted it as W. And define W_0, W_1 as Bayes Classifier is wrong and with a prediction 0, 1, alsp W as the Bayes Classifier is wrong. And R_0, R_1 as Bayes Classifier iw right and with a result 0, 1 repectly. Note the results of two classifiers are different with symbol D.

$$
\begin{aligned}
P(W) &\leq P(W|W_0)P(W_0) + P(W|W_1)P(W_1) + P(W|R_0)P(R_1)P(W|R_1)P(R_1) \\
&\leq P(W_0) + P(W_1) + P(W|R_0) + P(W|R_1) \\
&\leq 2P(W) + P(D)
\end{aligned}
$$

Then, calculate the expectation in the trainging set, one can derive

$$
E_S\mathcal{E}(f^{1NN}) \leq 2\mathcal{E}f^* + E_S E_x\|\eta - \eta_\pi\|
$$

As the case D is totally equivalent with the result of the absolute value.

Use the Lipschitz Condition, one can derive the result.