CIS 419/519: Applied Machine Learning — Fall 2021

Homework 1

Handed Out: September 15 — *Due: September 27*

Name: QIHANG DAI

PennKey: ahgdyycc

PennID: 78803164

1 Multiple Choice & Written Questions

1. (a) i. decrease the training error, cause the model is overfitting without regularization on x related value
   ii. increase the training error, infinite regularization, underfitting
   iii. increase the training error, same as 2

   (b) i. the intercept can be zero since two class are equal
   ii. class 1 have more possibility. θ_0 should be larger so exp(-θ_0) is smaller and the probability is larger

2. (a) since there is only two point, the boundary should be a perpendicular line to the line connecting two points

   (b) k = 1, so each data point is its own neighbor, for the dataset each data must have it own label thus the decision boundary is acheived

   (c) k = infinite, all data points are neighbors, the family would be a constant model that predict all the same output regardless of input

   (d) when k is inifinite, the bias is high cause underfitting. when k is 1, the variance is high cause overfit

   (e) instead of majority vote, we can use square distance, cubic distance, etc. to weight the vote. the higher order of the distance, it gives more weight on the closer points, which increase the true positive rate.

3. (a) see:

$$P(yes) = \frac{1}{2}$$

$$H(D) = -\frac{1}{2}log_2(\frac{1}{2}) - \frac{1}{2}log_2(\frac{1}{2}) = 1$$

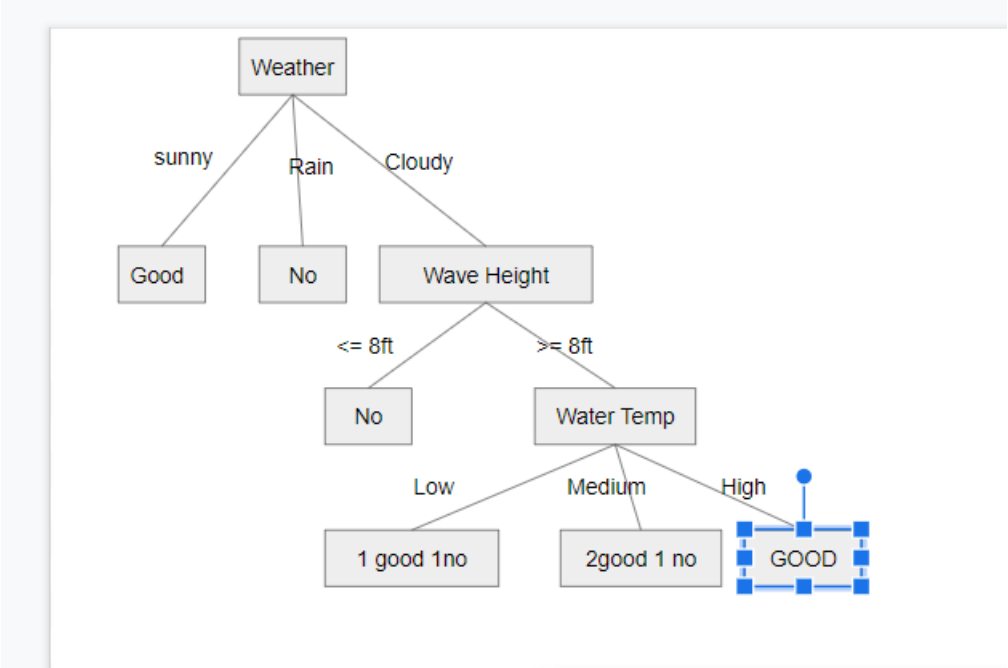
$$IG(D, Weather) = 1 - (\frac{3}{8} * (0) + \frac{2}{8} * (0) + \frac{3}{8}(-\frac{1}{3}log_2\frac{1}{3} - \frac{2}{3}log_2\frac{2}{3})) = 0.65$$

$$IG(D,WT) = 1 - (\frac{2}{8}(-\frac{1}{2}log_2\frac{1}{2} - \frac{1}{2}log_2\frac{1}{2}) \tag{1}$$
$$- \frac{3}{8}(-\frac{1}{3}log_2\frac{1}{3} - \frac{2}{3}log_2\frac{2}{3}) \tag{2}$$
$$- \frac{3}{8}(-\frac{1}{3}log_2\frac{1}{3} - \frac{2}{3}log_2\frac{2}{3})) \tag{3}$$
$$= \frac{6}{8}(1 - \frac{1}{3}log_2\frac{1}{3} - \frac{2}{3}log_2\frac{2}{3}) = 0.0675 \tag{4}$$

$$IG(D,Wh) = 1 - (\frac{3}{8}(0) + \frac{5}{8}(-\frac{1}{5}log_2(\frac{1}{5}) - \frac{4}{5}log_2(\frac{4}{5}))) = 0.55 \tag{5}$$

thus we choose Weather as the root node to split the data

(b) pic:



(c) the probability of be a good day to surf is 2/3, so should be a good day

(d) no. there is error in the leave node as shown in my tree

4. ans:
For real-valued input, we cant pick a set of thresholds to do binary split. Thus we can calculate different information gain based on different set of thresholds and pick the one with the highest information gain.
For the optimizer, along with the greedily choose the best IG, we can also publish the errorate of the node to gain a better performance, or use Gain ratio to avoid overfitting.

5. ans:

$$f_{\hat{\beta}}(x) = \hat{\beta}^T x = x^T \hat{\beta}$$
$$\hat{\beta} = (X^T X)^{-1} X^T Y$$
$$f_{\hat{\beta}}(x) = x^T (X^T X)^{-1} X^T Y$$
$$Y = (y1, y2, ..., yn)^T,$$
$$f_{\hat{\beta}}(x) = x^T (X^T X)^{-1} X^T (y1, y2, ..., yn)^T$$
$$= \sum_{i=1}^{n} x^T (X^T X)^{-1} X^T yi$$
$$k_i = x^T (X^T X)^{-1} X^T I_i$$

I_i represent (n x 1) vector where only ith element is 1 and others are 0.

2 Python Programming Questions

TODO: Place your figure and paragraph for Q2.2 here

TODO: Place your figure and paragraph for Q 3.1.2 here

TODO: Place your paragraph for Q3.2 here

TODO: Place your report for Q4.2 here

TODO: Place your paragraph for Q4.2.1 here

(if you are attempting 4.3, remember to include your confidence intervals in the performance table)