CSCI567TA1

Youzhi Qu USCID: 5045807938

September 2019

1 Problem 1 Nearest Neighbor Classification

Proof: if $C(\boldsymbol{x_i}, \boldsymbol{x_j}) \leq C(\boldsymbol{x_i}, \boldsymbol{x_o})$, then $E(\boldsymbol{x_i}, \boldsymbol{x_j}) \leq E(\boldsymbol{x_i}, \boldsymbol{x_o})$, where $\|\boldsymbol{x_i}\|_2 = \|\boldsymbol{x_j}\|_2 = \|\boldsymbol{x_o}\|_2 = 1$.

$$\begin{aligned}
E(\boldsymbol{x_i}, \boldsymbol{x_j}) &= \|\boldsymbol{x_i} - \boldsymbol{x_j}\|_2^2 \\
&= \textstyle\sum_{d=1}^{D} (x_{id} - x_{jd})^2 \\
&= \textstyle\sum_{d=1}^{D} (x_{id}^2 - 2 \cdot x_{id} \cdot x_{jd} + x_{jd}^2) \\
&= \textstyle\sum_{d=1}^{D} x_{id}^2 + \sum_{d=1}^{D} x_{jd}^2 - \sum_{d=1}^{D} 2 \cdot x_{id} \cdot x_{jd} \\
&= \|\boldsymbol{x_i}\|_2^2 + \|\boldsymbol{x_j}\|_2^2 - \textstyle\sum_{d=1}^{D} 2 \cdot x_{id} \cdot x_{jd} \\
&= 2 - 2 \cdot \textstyle\sum_{d=1}^{D} x_{id} \cdot x_{jd} \\
&= 2(1 - \textstyle\sum_{d=1}^{D} x_{id} \cdot x_{jd}) \\
&= 2(1 - \frac{\sum_{d=1}^{D} x_{id} \cdot x_{jd})}{\|\boldsymbol{x_i}\|_2 \|\boldsymbol{x_j}\|_2} \\
&= 2 \cdot C(\boldsymbol{x_i}, \boldsymbol{x_j})
\end{aligned}$$

if $C(\boldsymbol{x_i}, \boldsymbol{x_j}) \leq C(\boldsymbol{x_i}, \boldsymbol{x_o})$, then $2 \cdot C(\boldsymbol{x_i}, \boldsymbol{x_j}) \leq 2 \cdot C(\boldsymbol{x_i}, \boldsymbol{x_o})$.

According to $E(\boldsymbol{x_i}, \boldsymbol{x_j}) = 2 \cdot C(\boldsymbol{x_i}, \boldsymbol{x_j})$, therefore $E(\boldsymbol{x_i}, \boldsymbol{x_j}) \leq E(\boldsymbol{x_i}, \boldsymbol{x_o})$.

2 Problem 2 Nearest Neighbor Classification and Decision Trees

1. Can we have a decision tree to classify the dataset with zero classification error?

Yes. Because data $\boldsymbol{x}$ is 100 dimensional binary vector, we can use perfect binary tree with depth 100 to classify the dataset with zero classification error.

2. Can we specify a 1-NN to result in exactly the same classification as our decision tree?

Yes. Because K=1, we always choose the nearest point in the training dataset. when we use training dataset to test, we always get the point itself without error. Therefore, 1-NN can get zero traning error.

3 Problem 3 Nearest Neighbor Classification and Decision Trees

Yes, we can use Feature{{A+1,B+1},{A+1,B-1},{A-1,B+1},{A-1,B-1}}, Label{{class1},{class0},{class0},{class1}} to be 1-NN training dataset.

4 Problem 4 Decision Tree

4.1

test error is 0.

4.2

test error is $\frac{1}{2}$.

4.3

(1)yes (2)no

4.4

No. We can not get zero classification error by using depth-1 decision tree. linear combination of variables x_1 and x_2 can only divide plate into two parts not four parts which we need in order to get zero classification error.

5 Problem 5 Decision Trees

5.1

classification error:

τ_1 left leaf $= 0.25$

τ_1 right leaf $= 0.25$

τ_2 left leaf $= 0$

τ_2 right leaf $= 0.33$

entropy:

τ_1 left leaf: $-\frac{150}{200} \times \log_e \frac{150}{200} - \frac{50}{200} \times \log_e \frac{50}{200} = 0.56$

τ_1 right leaf: $-\frac{50}{200} \times \log_e \frac{50}{200} - \frac{150}{200} \times \log_e \frac{150}{200} = 0.56$

τ_2 left leaf: $-\frac{100}{100} \times \log_e \frac{100}{100} = 0$

τ_2 right leaf: $-\frac{100}{300} \times \log_e \frac{100}{300} - \frac{200}{300} \times \log_e \frac{200}{300} = 0.64$

Gini impurity

τ_1 left leaf: $\frac{150}{200} \times (1 - \frac{150}{200}) + \frac{50}{200} \times (1 - \frac{50}{200}) = 0.38$

τ_1 right leaf: $\frac{50}{200} \times (1 - \frac{50}{200}) + \frac{150}{200} \times (1 - \frac{150}{200}) = 0.38$

τ_2 left leaf: 0

τ_2 right leaf: $\frac{200}{300} \times (1 - \frac{200}{300}) + \frac{100}{300} \times (1 - \frac{100}{300}) = 0.44$

5.2

Classification error rate:

$\tau_1 : \frac{100}{400} = 0.25$

$\tau_2 : \frac{100}{400} = 0.25$

In terms of Classification error rate, τ_2 is equal to τ_1

Conditional entropy:

$\tau_1 : \frac{200}{400} \times 0.56 + \frac{200}{400} \times 0.56 = 0.56$

$\tau_2 : \frac{100}{400} \times 0 + \frac{300}{400} \times 0.64 = 0.48$

In terms of Conditional entropy, τ_2 is better than τ_1

Weighted Gini impurity:

$\tau_1 : \frac{200}{400} \times 0.38 + \frac{200}{400} \times 0.38 = 0.38$

$\tau_2 : \frac{100}{400} \times 0 + \frac{300}{400} \times 0.45 = 0.34$

In terms of Weighted Gini impurity, τ_2 is better than τ_1

6 Problem 6 Naive Bayes

6.1

$P(PlayTennis = Yes) = \frac{2}{3}$

$P(PlayTennis = No) = \frac{1}{3}$

6.2

$P(Weather = Sunny|PlayTennis = Yes) = \frac{1}{2}$

$P(Emotion = Normal|PlayTennis = Yes) = \frac{1}{4}$

$P(Homework = Much|PlayTennis = Yes) = \frac{1}{4}$

6.3

$P(Weather = Sunny|PlayTennis = No) = \frac{1}{2}$

$P(Emotion = Normal|PlayTennis = No) = \frac{1}{2}$

$P(Homework = Much|PlayTennis = No) = \frac{1}{2}$

$P(x|PlayTennis = Yes) = \frac{1}{32}$

$P(x|PlayTennis = No) = \frac{1}{8}$

$P(x) = \frac{1}{16}$

$$\begin{aligned}
P(PlayTennis = Yes|x) &= \frac{P(x|PlayTennis=Yes)P(PlayTennis=Yes)}{P(x)} \\
&= \frac{1}{3} \\
P(PlayTennis = No|x) &= \frac{P(x|PlayTennis=No)P(PlayTennis=No)}{P(x)} \\
&= \frac{2}{3}
\end{aligned}$$