Machine Learning, 2021 Fall
Quiz 2 solution

1 [Lecture 9-10]

Bias, measures the distance of the average response of the model from the true value, high bias implies this underfitting.

Variance: measures the generalizability of model beyond the training data, high variance implies overfitting.

Noise: the inaccuracy of data.

Expected error: the expected error between model and true mapping fuction.

2 [Lecture 11-12]

True, we can always rescale the parameters to make the margin = 1.

False, you need to pick penalty term C.

3 [Lecture 13]

We will use the dataset below to learn a decision tree which predicts if people pass machine learning (Yes or No), based on their previous GPA (High, Medium, or Low) and whether or not they studied.

GPA	Studied	Passed
L	F	F
L	T	T
M	F	F
M	T	T
H	F	T
H	T	T

For this problem, you can write your answers using $\log_2$, but it may be helpful to note that $\log_2 3 \approx 1.6$

1. What is the entropy H(Passed) ?
   ANSWER:
   $$H(\text{ Passed }) = -\left(\tfrac{2}{6}\log_2\tfrac{2}{6} + \tfrac{4}{6}\log_2\tfrac{4}{6}\right)$$
   $$H(\text{ Passed }) = -\left(\tfrac{1}{3}\log_2\tfrac{1}{3} + \tfrac{2}{3}\log_2\tfrac{2}{3}\right)$$
   $$H(\text{ Passed }) = \log_2 3 - \tfrac{2}{3} \approx 0.92$$
2. What is the entropy H(Passed | GPA) ?
   ANSWER:
   $$H(\text{ Passed } \mid GPA) = -\tfrac{1}{3}\left(\tfrac{1}{2}\log_2\tfrac{1}{2} + \tfrac{1}{2}\log_2\tfrac{1}{2}\right) - \tfrac{1}{3}\left(\tfrac{1}{2}\log_2\tfrac{1}{2} + \tfrac{1}{2}\log_2\tfrac{1}{2}\right) - \tfrac{1}{3}\left(1\log_2 1\right)$$
   $$H(\text{ Passed } \mid GPA) = \tfrac{1}{3}(1) + \tfrac{1}{3}(1) + \tfrac{1}{3}(0)$$
   $$H(\text{ Passed } \mid GPA) = \tfrac{2}{3} \approx 0.66$$

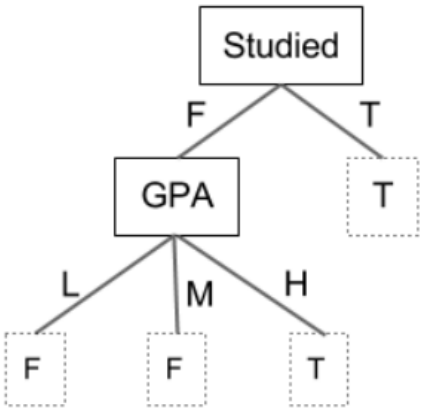
3. What is the entropy H(Passed | Studied) ?
ANSWER:

$$H(\text{ Passed } \mid \text{ Studied }) = -\frac{1}{2}\left(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right) - \frac{1}{2}\left(1\log_2 1\right)$$

$$H(\text{ Passed } \mid \text{ Studied }) = \frac{1}{2}\left(\log_2 3 - \frac{2}{3}\right.$$

$$H(\text{ Passed } \mid \text{ Studied }) = \frac{1}{2}\log_2 3 - \frac{1}{3} \approx 0.46$$

4. Draw the full decision tree that would be learned for this dataset. You do not need to show any calculations.
ANSWER: We want to split first on the variable which maximizes the information gain $H($ Passed $) - H($ Passed $\mid A)$. This is equivalent to minimizing $H($ Passed $\mid A)$, so we should split on "Studied" first.



4 [Lecture 14]

effective number is N/k, if the neighborhoods were nonoverlapping, there would be N/k classes and we would fit one parameter (a mean) in each class.

(If some students do not answer effective number, but analysis that "as k decreasing, model overfitting, larger variance and lower bias", they can recieve at most 0.2 points).

5 [Lecture 15]

(a) Cut(A, B): Minimize the weights of connections between groups

Assoc(A, B): Maximize the weights of connections within groups

Ncut(A, B): Consider the connectivity between groups relative to the density of each group

Nassoc(A, B): Consider the connectivity within groups relative to the density of each group

(b) Because they consider the intra-cluster density

6 [Lecture 16]

The left singular vectors measures document-concept similarity, the right singular vectors measures term-concept similarity.