Aaron Morgenegg, A02072659

CS5662 HW 5

Tan 4.6

a. 0.277
b. Undergraduate is more likely, because there is an 80% chance a student is an undergrad, vs 20% chance of a graduate student.
c. Undergrads are more likely, because probability a smoker is an undergrad is 0.723, while probability a smoker is a graduate is 0.277.
d. Graduate student is more likely.

Tan 4.7

P(A|+) = 0.6
P(B|+) = 0.2
P(C|+) = 0.4
P(A|-) = 0.4
P(B|-) = 0.4
P(C|-) = 1

Tan 4.10

Tan 4.11

a. 0.029
b. 0.002
c. 0.008

Tan 4.12

a.
1 nearest neighbour → +
3 nearest neighbour → -
5 nearest neighbour → +
9 nearest neighbour → -
b.
1 nearest neighbour → +
3 nearest neighbour → +
5 nearest neighbour → +
9 nearest neighbour → +

Tan 4.21

a.
decision tree: Would work well for this data set.
naive Bayes: Would work well for this data set.
k-nearest neighbour: Does not do well due to noise.
b.
decision tree: Would work well for this data set.

naive Bayes: Would work well for this data set.
k-nearest neighbour: Would work well for this data set.

c.

decision tree: Would work well for this data set.
naive Bayes: Would work well for this data set.
k-nearest neighbour: Would work well for this data set.

d.

decision tree: Would work well for this data set.
naive Bayes: Would work well for this data set.
k-nearest neighbour: Would work well for this data set.

e.

decision tree: Would work well for this data set.
naive Bayes: Would work well for this data set.
k-nearest neighbour: Would work well for this data set.

f.

decision tree: Would work well for this data set.
naive Bayes: Would work well for this data set.
k-nearest neighbour: Would work well for this data set.