

Tutorial 5

Bayesian Classifier In this task, we will be using the same weather dataset that was used in the tutorial on decision trees. The data is shown in Table 1.

outlook	temperature	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Table 1: The weather data (Witten and Frank; 1999, p. 9).

Instructions:

In this dataset, there are five categorical attributes *outlook*, *temperature*, *humidity*, *windy*, and *play*. We are interested in building a system which will enable us to decide whether or not to play the game on the basis of the weather conditions, *i.e.* we wish to classify the data into two classes, one where the attribute *play* has the value “yes”, and the other where it has the value “no”. This classification will be based on the values of the attributes *outlook*, *temperature*, *humidity*, and *windy*.

Question 1 Calculate $P(C1 | X = (\text{sunny}, \text{hot}, \text{high}, \text{false}))$. How would the Bayes classifier classify the data instance $X = (\text{sunny}, \text{hot}, \text{high}, \text{false})$?

Question 2 Does this agree with the classification given in Table 1 for the data instance

$X = (\text{sunny}, \text{hot}, \text{high}, \text{false})$?

Question 3 Consider a new data instance $X' = (\text{overcast}, \text{cool}, \text{high}, \text{true})$. How would the Bayes classifier classify X' ?

