Problem Set 2

Person	Rain Forecast (cm)	Sick	Attend Class?
p1	0	Yes	Yes
p2	0.2	No	Yes
p3	1	No	Yes
p4	2	Yes	No
p5	3	No	No
p6	0.7	Yes	No
p7	0.4	Yes	Yes
p8	0.4	No	Yes

Table 1: Training Data.

Person	Rain Forecast (cm)	Sick	Attend Class?
p9	1.5	No	Yes
p10	1.7	Yes	No
p11	0.5	No	Yes
p12	0.8	Yes	No

Table 2: Test Data.

1. **Decision Tree (10 points)**:

   (a) Show the decision tree that results from using the training data shown in Table 1 with the entropy measure to split features. For full credit, you must show the key mathematical steps towards developing your tree.

   (b) Report prediction results for the test data shown in Table 2.

   (c) Evaluate the prediction results by reporting the precision and recall.

   (d) Evaluate the prediction results by showing a confusion matrix.

2. **Naive Bayes (10 points)**:

   (a) Show the result of training a Gaussian Naive Bayes model using the training data shown in Table 1, with only the "Snow Forecast" feature (ignore the "Sick" feature). Show the key mathematical steps to derive all class probabilities (i.e., $P(C_i)$) and likelihoods of features given the two classes (i.e., $P(feature|C_i)$).

   (b) Report prediction results for the test data shown in Table 2.

   (c) Evaluate the prediction results by reporting the precision and recall.

   (d) Evaluate the prediction results by showing a confusion matrix.

3. **Classification Models (5 points)**: In your own words, describe the difference between a generative model and a discriminative model