

1. [4 Points]: Explain what is the bias-variance trade-off? Describe a few techniques to reduce bias and variance respectively.

The bias-variance tradeoff is the trade-off that occurs when determining the number of parameters to use in a model. If there are only a few parameters, then there may be a large amount of bias even though there's a low amount of variance. If the model has a lot more parameters, then there may be a high amount of variance, even though there is a small amount of bias. A way to reduce bias is to increase the complexity of the model, and a way to reduce variance is to conduct resampling, such as through the random forest method.

2. [6 points] Assume the following confusion matrix of a classifier. Please compute its
 - 1) Precision,
 - 2) Recall,
 - 3) F1-Score

Predicted Results

	Class 1	Class 2
Class 1	50	30
Class 2	40	60

Precision

The precision is calculated by dividing the actual results of a class by the total number of predicted results for that class.

The precision for class 1 is $50/90 = 0.556$.

The precision for class 2 is $60/90 = 0.667$.

Recall

The recall is calculated by dividing the actual results by the total of actual results.

The recall for class 1 is $50/80 = 0.625$

The recall for class 2 is $60/100 = 0.6$

F1-Score

The F1-score is calculated by dividing the product of the precision and recall of a class by the sum of the precision and recall for a class, multiplied by 2.

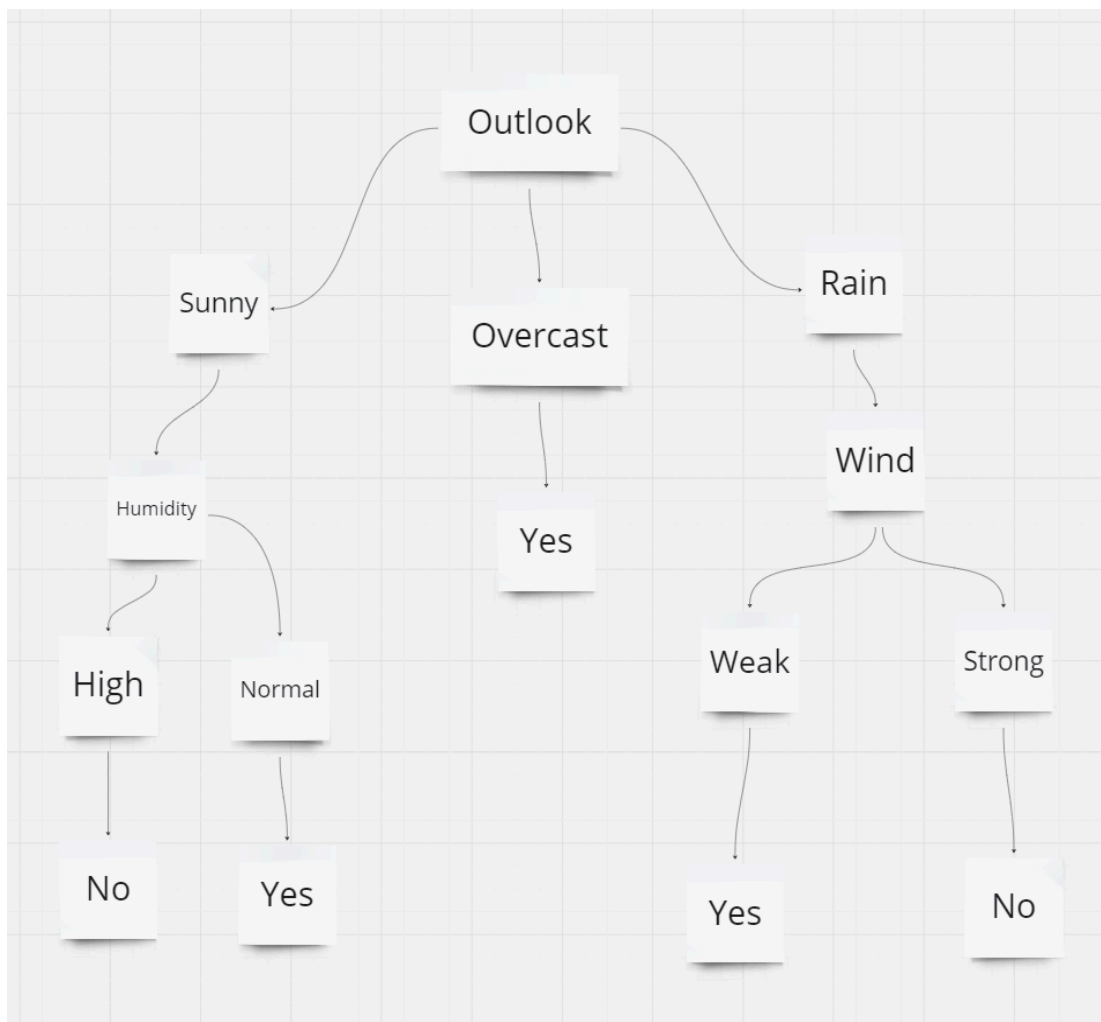
The F1 score for class 1 is $2((0.56 \times 0.625) / (0.56 + 0.625)) = 2(0.35 / 1.185) = 0.591$.

The F1 score for class 2 is $2((0.667 \times 0.6) / (0.667 + 0.6)) = 2(0.402 / 1.267) = 0.635$.

3. [10 points] Build a decision tree using the following training instances (using information gain approach):

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Weak	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes

Here is my decision Tree:



4. [10 Points] The naive Bayes method is an ensemble method as we learned in Module 5. Assuming we have 3 classifiers, and their predicted results are given in table 1. The confusion matrix of each classifier is given in table 2. Please give the final decision using the Naive Bayes method:

Table 1 Predicted Results of each Classifier:

Sample x	Result
Classifier 1	Class 1
Classifier 2	Class 1
Classifier 3	Class 2

Table 2 Confusion matrix of each classifier:

	i) Classifier 1			ii) Classifier 2			iii) Classifier 3	
	Class 1	Class 2		Class 1	Class 2		Class 1	Class 2
Class 1	40	10	Class 1	20	30	Class 1	50	0
Class 2	30	20	Class 2	20	30	Class 2	40	10

Class 1

Classifier 1: $40/70 = 0.571$

Classifier 2: $20/40 = 0.5$

Classifier 3: $0/10 = 0$

Class 2

Classifier 1: $30/70$

Classifier 2: $20/40$

Classifier 3: $10/10$

Class 1: 0

Class 2: 0.214

Result: Class 2