1. Explain what is the bias-variance trade off? Describe few techniques to reduce bias and variance respectively.

Bias/Variance Tradeoff is the theoretical idea that a model error can be expressed as a sum of 3 different errors.

1. The Bias which is defined as part of the generalization error which happens due to the wrong assumptions, such as assuming that the data is linear when in fact the data is quadratic. A high bias model is most likely to underfit the training data.
2. The Variance which happens when the model is excessively sensitive to a small variation in the training data. A model that has many degrees of freedom this can happen when a model has a high degree polynomial this will generally have a high variance, and thus can overfit the training data.
3. Lastly the Irreducible error this happens generally due to noisiness of the data. The only way to reduce this part of the error is to clean up the data. For example, fix any data sources such as sensors, or detect and remove outliers.

One way to increase the variances but decrease bias is by increasing the model’s complexity. On the other hand, reducing a model’s complexity will do just the opposite, this will increase the bias but reduce variance. This is exactly why this is called the bias/variance tradeoff. The below photo shows a perfect depiction of the Bias Variance Tradeoff. For High Variance regardless of if it is Low Bias or high Bias our sample data is more scattered, with Low Variance our data is closer together, in Low Bias we are right in the bullseye, and with High Bias we migrate further away from the bullseye.

There are a few techniques that can aid with fixing high bias, as well as High Variance. Fixing High Bias can be accomplished by 3 main techniques.

1. Adding more input features this will improve the data to fit better.
2. Add more polynomial features to improve the complexity of the model
3. Decrease the regularization term to have a balance between bias and variance.

On the other hand, there are also 2 ways to fix High Variance.

1. Reduce the input features, use only features with more feature importance to reduce overfitting the data.
2. Getting more training data will also help reducing high variance. This is because a high variance model will not be working for an independent dataset if you have very data.

Shape, circle

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

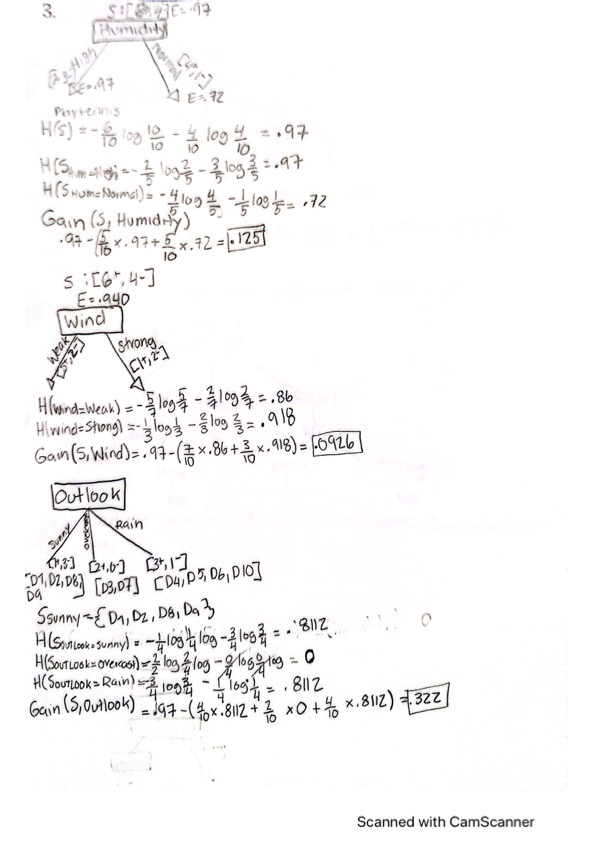
**Precision** TP/TP+FP = 50/ 50+40 =0.55556

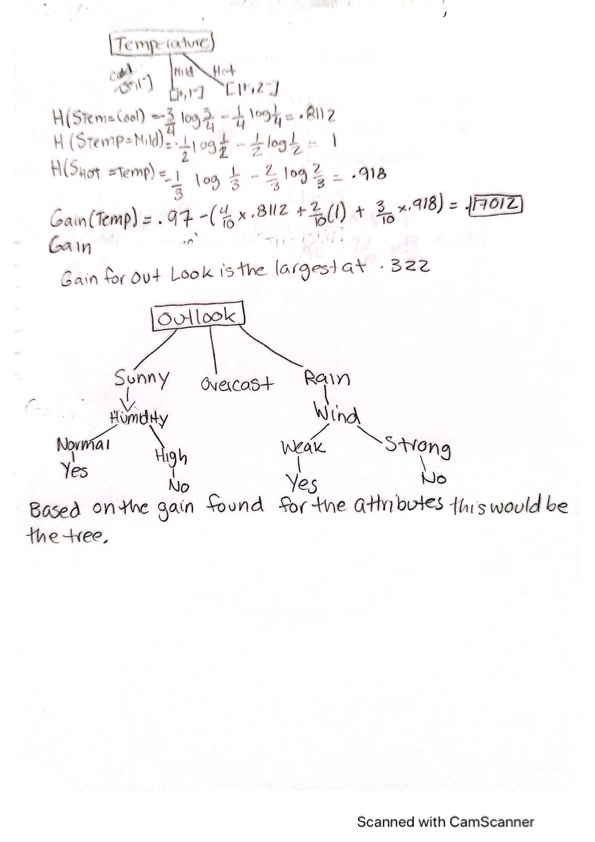
**Recall** TP/TP+FN = 50/50+30 = 0.625

**F1** 2pr /p+r = 2(.55556) x(.625)/ .55556+ .625 =.6944444/1.181=0.5880

Table

Description automatically generated





4.)

W1= represents the true class of the data(row in the confusion matrix)

D=two subscripts, first=classfier, second= output of individual classifier(1,1,2)

