

## Homework Problem Set 5: Ensemble methods, imbalanced classes

Due Wednesday, March 22 at 11:59 PM

Upload a pdf to Canvas

Each question is worth the same number of points.

---

### Question 1:

Given the Bayesian network shown in Figure 1 compute the following probabilities:

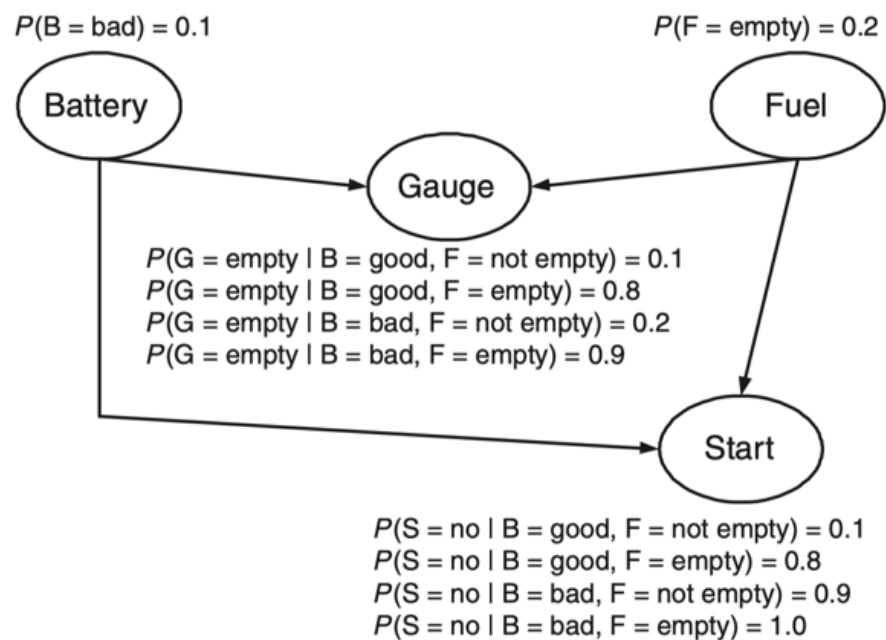


Figure 1: Bayesian Network

- (a)  $P(B = \text{good}, F = \text{empty}, G = \text{empty}, S = \text{yes})$ .
- (b)  $P(B = \text{bad}, F = \text{empty}, G = \text{notempty}, S = \text{no})$ .
- (c) Given that the battery is bad, compute the probability that the car will start.

---

### Question 2:

Boosting and bagging are ensemble methods used to improve the accuracy of predictive models. Below are questions divided into conceptual understanding and calculations to demonstrate the differences between boosting and bagging and the potential issues with boosting.

- (a) **Select *all* the correct statements related to boosting improvement strategies:**
  - (A) Boosting reduces bias by sequentially focusing more on the instances that previous models misclassified.

- (B) Boosting employs parallel model training to reduce computational time.
- (C) Boosting increases model diversity by training each model on a random subset of the dataset.
- (D) Boosting assigns weights to each training instance, adjusting these weights after each round to prioritize harder-to-classify instances in subsequent models.

(b) **Select *all* the correct statements:**

- (A) Boosting can significantly increase the model's variance due to the high weight it may place on outliers.
- (B) Boosting often leads to underfitting, as it places too much emphasis on instances that are easy to classify.
- (C) The sequential nature of boosting can lead to increased training times compared to bagging.
- (D) Boosting's focus on reducing bias makes it less susceptible to overfitting compared to other ensemble methods.

(c) **Weight Update Calculation:** Suppose after the first round of boosting, a model achieves an error rate of 0.3 (30% error) on its training dataset. Calculate the weight update factor ( $\alpha$ ) for the next round using the formula found in the slides.

(d) **Model Weight Influence:** Given that an instance was misclassified in the previous round, and the calculated weight update factor  $\alpha$  from question 1 is used, calculate the new weight for this instance if its original weight was 1. The updated weight is calculated using  $new\_weight = original\_weight \times e^\alpha$ .

### Question 3:

Alan recently came across the concept of ensemble methods in a data mining class and decided to apply it to stock market prediction. In order to predict whether the stock market will rise or fall on a given day, he decided to flip a coin 1000 times and predicted the stock market to go up if heads turn up in majority and vice-versa. He thinks that this approach could get him a better prediction of the stock market because an ensemble of independent classifiers could potentially obtain a better prediction. Do you agree with him? Give a brief justification.

### Question 4:

For a binary classification problem, you are given a collection of base classifiers where every base classifier has an error rate of  $e$ . The ensemble prediction is simply the majority vote of the predictions of the base classifiers. State whether the following statements are true or false. No justification required.

- a) If  $e = 0.5$  and the predictions of all base classifiers are independent, the error rate of the ensemble classifier will be smaller than  $e$ .
- b) If  $e = 0.3$  and the predictions of all base classifiers are independent, the error rate of the ensemble classifier will be smaller than  $e$ .

- c) If  $e = 0.3$  and all base classifiers are identical, the error rate of the ensemble classifier will be smaller than  $e$ .

---

**Question 5:**

On the left, you are given four classification scenarios that are possible for different settings of bias and variance of the classifier. On the right, you are given four types of classifiers. You need to match the classification scenarios to their corresponding classifiers either by drawing arrows or by writing down matching pairs. —

(a) Low Bias, Low Variance	(i) Underfitting Classifier
(b) High Bias, Low Variance	(ii) Overfitting Classifier
(c) Low Bias, High Variance	(iii) Ideal Classifier
(d) High Bias, High Variance	(iv) Worst Classifier

Table 1: Example Table

---

**Question 6:**

Consider a test data of 1000 samples with two classes: + class (100 samples) and - class (900 samples). We have two random classifiers C1 and C2. Classifier C1 classifies test data to + class randomly with a probability  $p$  and classifier C2 classifies test data to + class randomly with a probability  $2p$ .

- What is the expected TPR and FPR for C1 and C2?
- Is C2 a better classifier than C1? Hint: The random guess line in an ROC curve corresponds to  $\text{TPR}=\text{FPR}$ .
- The expected precision for both C1 and C2 is  $1/10$ . Expected recall for C2 is twice that of C1 ( $2p$  and  $p$ , respectively). If we use precision and recall as the evaluation metrics, C2 appears to be a better classifier than C1. Which evaluation metric pair between {TPR and FPR} and {precision and recall} do you think is correctly indicating the relative performance of C2 and C1?

---

**Question 7:**

Consider a test data of 200 samples with two classes (+ and -). The following are two confusion matrices for Classifier C1 and C2

	Predicted class: +	Predicted class: -
Actual class: +	10	90
Actual class: -	10	90

Table 2: C1: Precision = 50%; Recall/TPR = 10%; F1-measure = 16.67%; FPR = 10%

- Based on the information given above, comment on the relative performance of these two classifiers.
- Which evaluation metric pair between {TPR and FPR} and {precision, recall, F1-measure} do you think is correctly indicating the relative performance of classifiers C1 and C2? Explain briefly.

	Predicted class: +	Predicted class: -
Actual class: +	50	50
Actual class: -	50	50

Table 3: C2: Precision = 50%; Recall/TPR = 50%; F1-measure = 50%; FPR = 50%

- (iii) Consider a new Classifier, C3, with the following confusion matrix on the same test data. Given the choice of C1, C2 and C3, which classifier would you prefer over the other two?

	Predicted class: +	Predicted class: -
Actual class: +	20	80
Actual class: -	10	90

Table 4: Precision = 66.67%; Recall = 20%; F1-measure = 30.7%; FPR = 10%

Explain briefly.

#### Question 8:

We have test data of 1000 samples with two classes: a + class (100 samples) and a - class (900 samples). Consider a random classifier C0 that classifies a test data instance to the + class randomly with a probability  $p$ .

- What is the expected precision and recall for C0?
- Write the expression for the F-measure of C0?
- Consider another classifier C1 whose F-measure is known to be 0.15. Is C1 better than a random classifier?

#### Question 9:

You are given the task of predicting whether there will be rain or sunshine tomorrow in Seattle. Results of your favorite classification algorithm on a test set of 1000 days are shown in the confusion matrix below.

	Predicted: Sunshine (+)	Predicted: Rain (-)
Actual: Sunshine (+)	80	70
Actual: Rain (-)	50	800

Table 5: Example Table

- Compute the precision, recall, F-measure, and accuracy for the confusion matrix. (Compute precision, recall, and the F-measure with respect to + class only.)
- Which of these metrics is a poor indicator of the overall performance of your algorithm? Which of these metrics is the best indicator of the overall performance? Give a one sentence reason why this is the case?

**Question 10:**

You are trying to evaluate two different blood tests, T1 and T2, that have been developed to detect a particular type of cancer. T1 had been evaluated on a population of 200 subjects, out of which 100 were known to be suffering from cancer, while the remaining 100 were healthy. T2 had been evaluated on a different population of 1000 subjects, out of which 100 were known to be suffering from cancer, while the remaining 900 were healthy. The results of these tests are shown in the following confusion matrices, along with the values of the following evaluation measures: TPR, FPR, Precision, the F-measure, and TPR/FPR.

Table 6: TPR: 0.4; FPR: 0.1, Precision: 0.8, F1-Score: 0.533, TPR/FPR: 4,

Dataset: (100 patients)	Predicted by Blood Test	
Actual	Cancer (+ class)	No Cancer (- class)
Cancer (+ class)	40	60
Cancer (- class)	10	90

Table 7: TPR: 0.4; FPR: 0.055, Precision: 0.44, F1-Score: 0.41, TPR/FPR: 7.2,

Dataset: (100 patients)	Predicted by Blood Test	
Actual	Cancer (+ class)	No Cancer (- class)
Cancer (+ class)	40	60
Cancer (- class)	50	850

- According to the F-measure, which test is better?
- According to TPR/FPR, which test is better?
- For this situation, which evaluation measure (between F1-Score and TPR/FPR) should you use to make your selection between the two tests, T1 or T2? Why?
- Give an example scenario where you would reverse your choice of evaluation measure that you made in part (b)? (That is, if you chose TPR/FPR in part (c), give an example of a scenario where you would prefer the F measure over TPR/FPR and vice-versa)