1. What is the "prior probability" in a Bayesian classifier? (6%) (SC)

(a) The probability of observing a particular feature
(b) The probability of a specific class in the dataset
(c) The probability of classifying an observation correctly
(d) The probability of a feature in the dataset
(e) The probability of an observation given the class.

Ans: (b)

2. Which of the following statement(s) is(are) true about Naïve Bayes? (10%) (MC)

(a) Due to its simplicity and efficiency, Naïve Bayes can provide quick training and prediction times, making it suitable for real-time or near-real-time applications.
(b) Naïve Bayes often performs well in high-dimensional datasets, where the number of features is much larger than the number of data points.
(c) The independence assumption underlying Naïve Bayes implies that the values of each attribute follow a Gaussian distribution.
(d) Naïve Bayes can deal with the dataset having both discrete and continuous attributes
(e) When using Naïve Bayes, all features are equally important and contribute independently to the final outcome.

Ans: (a)(b)(d)(e)

3. Rachel loves cats and dogs. She recorded some information of cats and dogs she encountered these days, containing their appearance and whether we could pet them. Using a naïve Bayesian assumption, what is the best possible outcome when we encounter a fat white dog which doesn't look sleepy. (X'= (Animal = dog, Body Type = fat, Looks sleepy = No, Color = white)) (CA)

A. Please provide the answer ratio of $\frac{P(Yes|X')}{P(No|X')}$, where "Yes" means we could pet it. (Estimate to the second decimal place. e.g., 0.01). (5%)
B. Predict whether we could pet it ("Yes" or "No"). (5%)

	Animal	Body type	Looks sleepy	Color	Could we pet them
1	cat	fat	Yes	Orange	No
2	dog	thin	No	Black	Yes
3	cat	average	No	White	Yes
4	dog	average	Yes	Black	Yes
5	dog	fat	Yes	Orange	Yes
6	dog	thin	Yes	White	Yes
7	cat	thin	No	White	No
8	dog	healthy	Yes	Black	No

Ans: 3.46, Yes

4. Which of the following statement(s) is(are) true about decision trees and random forests? (12%) (MC)

(a) Decreasing the maximum depth of the decision tree will regularize the model and reduce the risk of overfitting.
(b) The problem of “no attributes left” during tree construction “only” happens when there is noise in the dataset (Noise means we can find more than one instance with exact same features but different labels).
(c) In a decision tree, the same attribute with continuous values could be used more than once under the same branch.
(d) The name "Random Forest" derives from the selection of decision trees in a "random manner" when combining them into a "forest".
(e) In a random forest, the input dimensions of each decision tree could be different.
(f) Compared to Decision trees, random forests make the decision making process more robust to missing data.

Ans: (a)(b)(c)(e)(f)

5. Which ensemble technique is best suited for handling class imbalance issues (e.g. the "positive" class has very few examples compared to the "negative" class.) in classification tasks? (10%) (SC)
(a) Bagging
(b) Boosting
(c) Stacking
(d) None of the above

Ans: (b)

6. Given the instances in the table below, we want to build a decision tree to decide whether David would be late for school. If we split data into three groups with the feature "weather", what are the values of the **entropy before split** (4%), the **entropy after split** (the weighted sum) (4%), the **information gain** (4%)**,** and the **gain-ratio** (4%)**?** (Round the number to the third decimal place. e.g., 0.001) (CA)
Note: Please calculate entropy with log function base 2.

day	weather	stayed up late	late for school
1	cold	Yes	0
2	cold	No	1
3	cold	Yes	1
4	cold	No	1
5	sunny	Yes	0
6	sunny	Yes	1
7	sunny	No	1
8	rainy	Yes	0
9	rainy	No	1
10	rainy	No	0
11	cold	No	0

12	sunny	Yes	1
13	sunny	No	0
14	cold	No	1
15	cold	No	0

Ans: 0.997, 0.967, 0.030, 0.020

7. Decision trees can also deal with continuous input variables. Suppose we want to split the data below into the left subtree and right subtree. Find the best feature and threshold to split data. (6%) (CA)

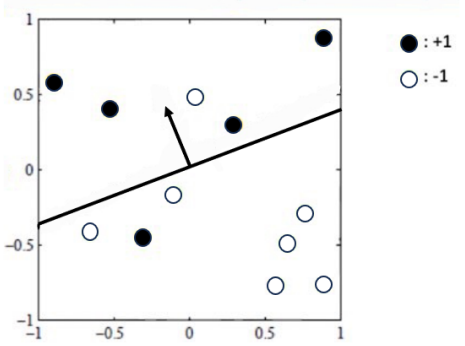
patient_id	age	BMI	diabetes_mellitus
1	18	38.830	1
2	39	20.830	0
3	60	22.722	1
4	73	39.409	1

Ans: BMI, 21.776

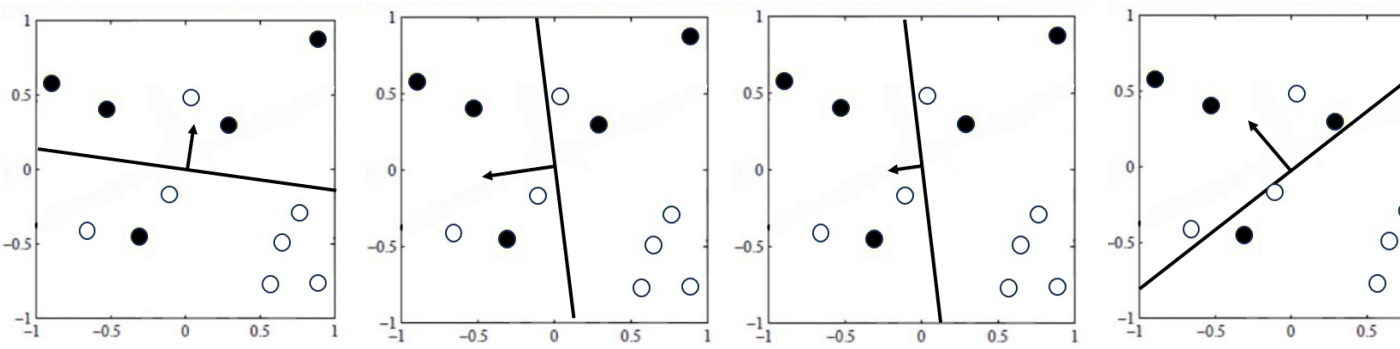
8. Which of the following statement(s) is(are) true about Bagging and Boosting? (10%) (MC)
   (a) In Boosting, base classifiers that have more wrongly classified cases get higher weights when combining all base classifiers into a single powerful classifier.
   (b) Boosting is a sequential process, the final classifier comes from one of the best classifiers in all rounds
   (c) In Boosting, the final classifier is always the one from the last round
   (d) The key principle behind creating multiple base models in ensemble techniques is to Ensure that base models have similar predictions.
   (e) In the Bagging learning algorithm, we attempt to train several weak learners with reduced dependency.

Ans: (e)

9. What is(are) the possible plot(s) for the next step in the perceptron algorithm? (10%) (MC)



(a) (b) (c) (d)



Ans: (b)(c)

10. Here are the questions for the hw2:
10.1 If we don’t need to generate pickle in the advanced part, which line can be removed? please answer the line number(e.g. 1) (5%) (SC)

```
1  training_df = pd.read_csv('hw2_advanced_training.csv')
2  testing_df = pd.read_csv('hw2_advanced_testing.csv')
3  y = training_df['diabetes_mellitus']
4  X = training_df.drop('diabetes_mellitus', axis=1)
5  # after implement GaussianNaiveBayesian (you can't answer this line)
6  gnb_classifier = GaussianNaiveBayesian()
7  gnb_classifier.build_table(X, y)
8  X_train, X_val, y_train, y_val = train_val_split(X, y, val_size=0.2, random_state=0)
9  gnb_classifier.build_table(X_train, y_train)
10 predictions = gnb_classifier.predict(X_train)
```

Ans: 7

10.2 Given that various features exhibit dissimilar value ranges and units, when considering standardization of features in the code, which option is the most appropriate answer “if the goal is to improve f1 score”? (Note: the Gaussian likelihood using the formula of the Gaussian probability density function (pdf) as described below) (5%) (SC)

```
1  def standardization(df):
2      for column in df.columns:
3          mean = df[column].mean()
4          std = df[column].std()
5          df[column] = (df[column] - mean) / std
6
```

```
training_df = pd.read_csv('hw2_advanced_training.csv')
testing_df = pd.read_csv('hw2_advanced_testing.csv')
y = training_df['diabetes_mellitus']
X = training_df.drop('diabetes_mellitus', axis=1)

# after implement GaussianNaiveBayesian

# standardization(X)                  # Question 10.2 option A
# standardization(testing_df)         # Question 10.2 option A

gnb_classifier = GaussianNaiveBayesian()
gnb_classifier.build_table(X, y)

# standardization(X)                  # Question 10.2 option B
# standardization(testing_df)         # Question 10.2 option B

# after generate pickle file

X_train, X_val, y_train, y_val = train_val_split(X, y, val_size=0.2, random_state=0)

gnb_classifier = GaussianNaiveBayesian()
gnb_classifier.build_table(X_train, y_train)
predictions = gnb_classifier.predict(X_train)
```

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

(a) Option A may enhance the f1-score of predictions
(b) Option B may enhance the f1-score of predictions
(c) Either choosing A or B might enhance the f1-score of predictions
(d) Either choosing A or B is unlikely to affect the f1-score

Ans: (d)