STAT 451 Final Exam

Question	Points	Earned
Q1 (cover)	1	
Q2	8	
Q3	8	
Q4	8	
Q5	8	
Q6	18	
Q7	8	
Q8	8	
Q9	8	
Total	75	

1. If a question is ambiguous, resolve the ambiguity in writing. We will consider grading accordingly. e.g.

   - In #10, I think "average" refers to the population mean μ (not the sample mean $\bar{X}$).
   - In #13b, I think ...

   Please answer this question with a period (.) if you have no other comment, so that Canvas will think you answered it and give you its 1 point. Do not write unnecessary comments.

Answer each question at

https://canvas.wisc.edu/courses/355816/quizzes/460845 **as you work through the exam** so that you do not run out of time with questions unanswered.

2. Consider using k-means on the unsupervised 1D dataset $\{\mathbf{x}\} = \{1, 3, 5, 10, 12\}$ to create $k = 2$ clusters. Suppose the two initial randomly-chosen cluster centroids are $\mathbf{c}_1 = 3$ and $\mathbf{c}_2 = 5$.

   (a) What are the centroids after the first iteration of k-means?
   $\mathbf{c}_1 =$ _____ and $\mathbf{c}_2 =$ _____.

   (b) What are the centroids after the second iteration?
   $\mathbf{c}_1 =$ _____ and $\mathbf{c}_2 =$ _____.

3. For each situation, indicate which hyperparameter search strategy, G = grid search or R = random search, is more likely to be successful. Suppose computation time is limited.

   (a) _____ A model has two hyperparameters. The first takes one of two string values and the other takes one of three numeric values.

   (b) _____ A model has two hyperparameters. The first takes a floating-point number in the interval $[0, 1]$ while the second takes an integer in the range $[0, 100000]$.

4. Consider the use of bagging applied to classification decision trees of depth 1 (one decision node and two leaf nodes per tree). A training data set, on the left, consists of $\{(\mathbf{x}, y)\} = \{(x, y)\}$ because $\mathbf{x}$ has only one feature, x. It is followed by $B = 3$ bootstrap resamples created by sampling with replacement from the training data.

Training data

x	y
1	0
2	1
3	0
4	1

Resample #1

x	y
1	0
2	1
4	1
4	1

Resample #2

x	y
1	0
1	0
3	0
4	1

Resample #3

x	y
1	0
1	0
2	1
2	1

Consider making a prediction for $\mathbf{x} = 2$.

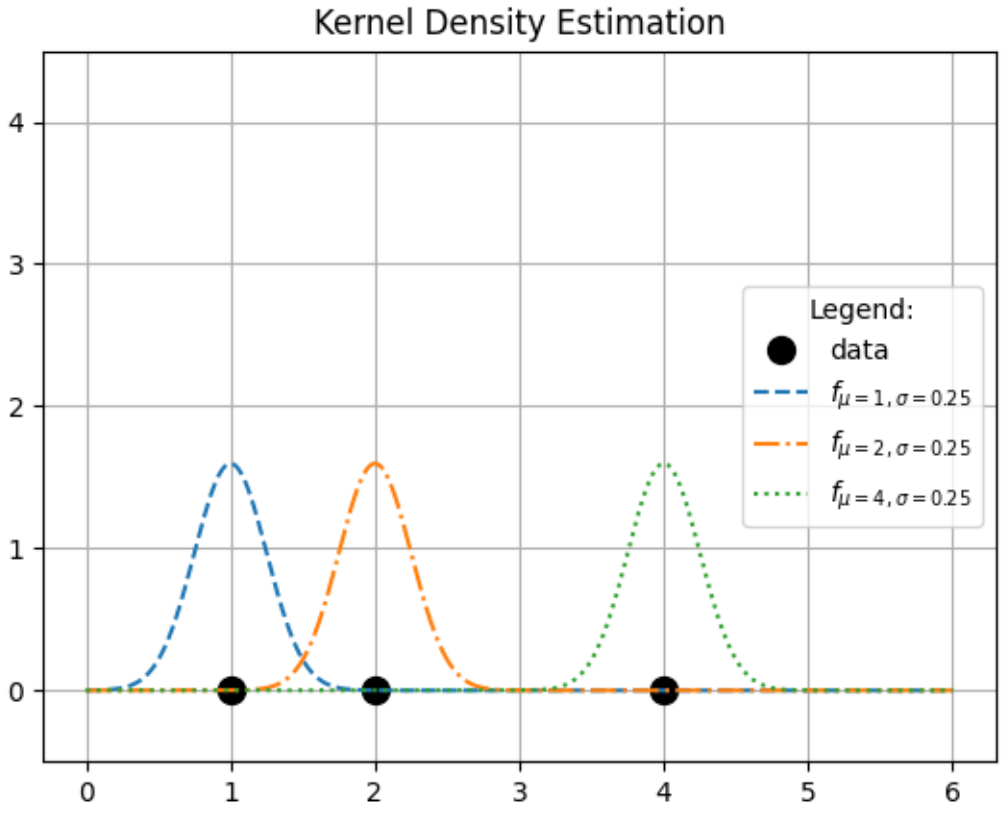
(a) What prediction is made by the tree trained on Resample #1? $\hat{y} =$ _____

(b) What prediction is made by the tree trained on Resample #2? $\hat{y} =$ _____

(c) What prediction is made by the tree trained on Resample #3? $\hat{y} =$ _____

(d) What prediction is made by this bagging classifier? $\hat{y} =$ _____

5. Here is a graph of 1D data $\{\mathbf{x}_i\} = \{x_i\} = \{1, 2, 4\}$ and corresponding Gaussian curves $\{f_{\mu=x_i,\sigma=b}(x)\}$ made with bandwidth $b = 0.25$.



(a) Supposing the data were randomly sampled from some population, use kernel density estimation to estimate the population's probability density $f(x)$ at $x = 1$.
Based on the plot, the estimate is $\hat{f}_{b=0.25}(1) \approx$ _____.

(b) Estimate the density at $x = 1.5$.
Based on the plot, the estimate is $\hat{f}_{b=0.25}(1.5) \approx$ _____.

(c) On the figure above, draw the estimated density function over the interval $[0, 6]$.

6. Consider the following questions about model assessment.

(a) Consider a classifier trained on examples $(\mathbf{x}, y)$ in the first two columns of the table below that makes the predictions on training data in the third column.

$\mathbf{x}$	y	$\hat{y}$
(0, 0)	0	0
(1, 4)	1	1
(3, −2)	1	1
(3, 0)	0	1

Complete the corresponding confusion matrix:

	predicted $\hat{y}$	
actual y	0	1
0		
1		

(b) The classifier is evaluated on unseen test data yielding this confusion matrix:

	predicted $\hat{y}$	
actual y	0	1
0	2	4
1	3	5

What is the precision on the test data?

(c) What is the recall on the test data?

(d) What is the accuracy on the test data?

(e) For a classifier that is randomly guessing with $P(\hat{y} = 1) = \frac{1}{3}$, what is the AUC?

(f) For a classifier with TPR = 1 and FPR = 0, what is the AUC?

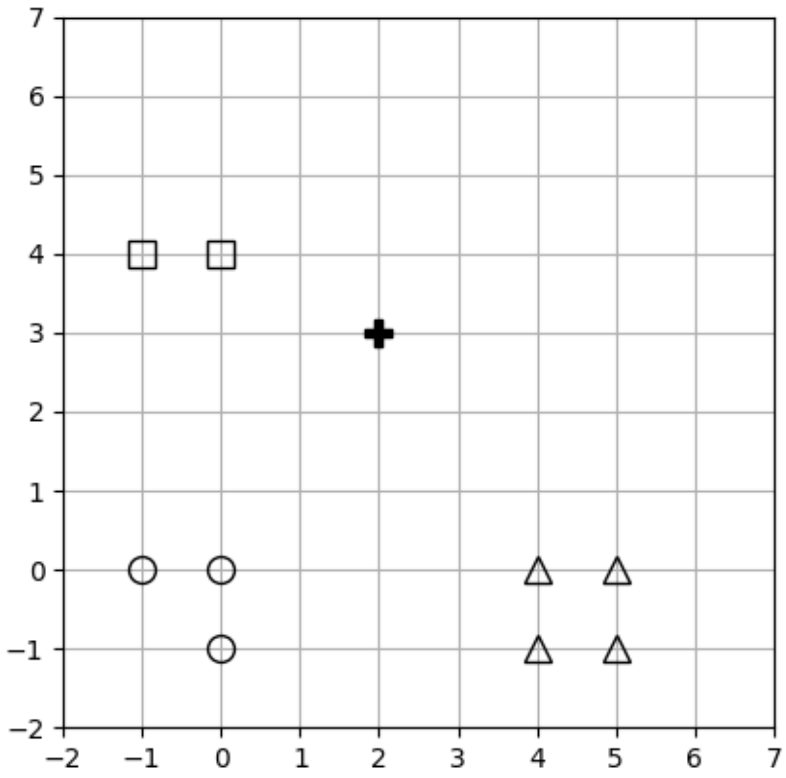
(g) For each situation, indicate whether P = precision or R = recall should be optimized:

i. _____ A bank is doing fraud detection where a fraudulent transaction ("positive") that is missed is expensive but a valid transaction labeled fraudulent is inexpensive.

ii. _____ A doctor is screening patients for a disease in which an ill patient ("positive") infects others and dies if the disease is not diagnosed.

iii. _____ A marketing campaign invests considerable expense in a prospective customer when it classifies that customer as likely to make a purchase ("positive").

7. Consider a one-vs.-rest SVM classifier trained on the following data depicted by circles, squares, and triangles:



(a) On the graph above, draw the three binary classifiers required by this method.

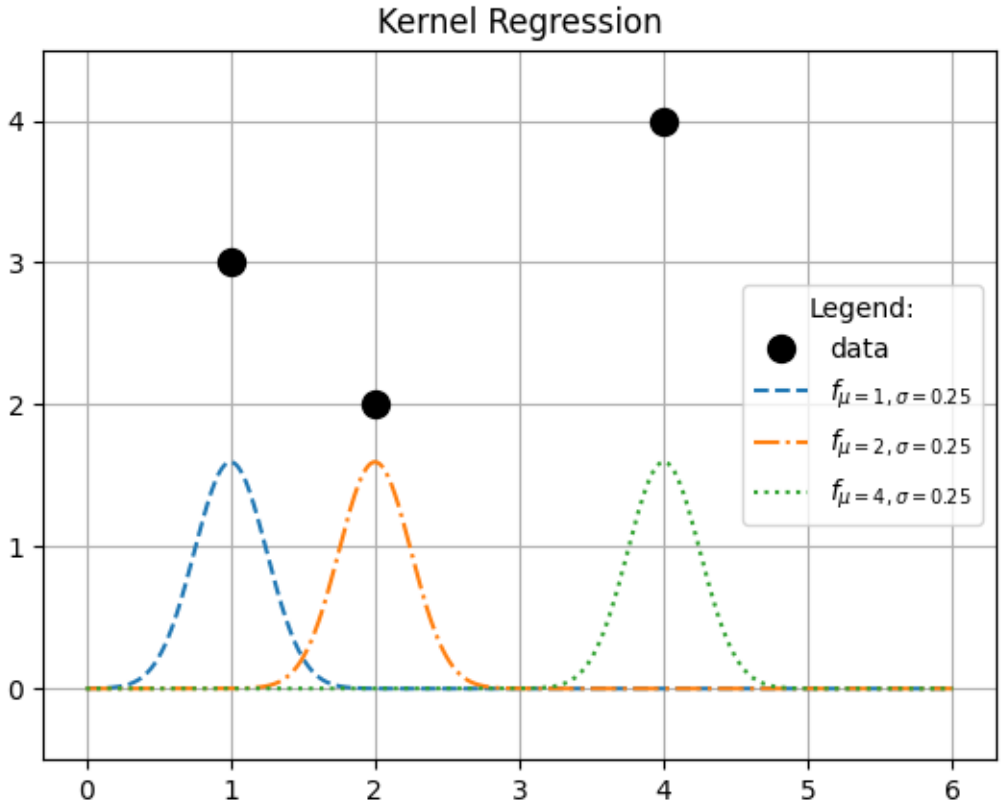
(b) How does this classifier classify the point indicated by "+"?

_____ circle
_____ square
_____ triangle

(c) Which category is ranked second by this classifier's decision method for the "+"?

_____ circle
_____ square
_____ triangle

8. Here is a graph of the data set $\{(\mathbf{x}_i, y_i)\} = \{(x_i, y_i)\} = \{(1,3),(2,2),(4,4)\}$ (here each $\mathbf{x}_i$ is a 1D x_i) along with corresponding Gaussian curves $\{f_{\mu=x_i,\sigma=b}(x)\}$ made with bandwidth $b = 0.25$:



(a) Use kernel regression to estimate $y = f(x)$ for $x = 1$.
Based on the plot, the estimate is $\hat{y} \approx$ _____.

(b) Estimate $y = f(x)$ for $x = 1.5$.
Based on the plot, the estimate is $\hat{y} \approx$ _____.

(c) On the figure above, draw the estimated regression function over the interval $[0, 6]$.

9. The next two questions are about principal component analysis (PCA).

(a) Consider the following code and its output:

```
rng = np.random.default_rng(seed=0)
(n_rows, n_cols) = (10, 4)
X = rng.normal(loc=0, scale=1, size=n_rows*n_cols).reshape((n_rows, n_cols))
pca = PCA(n_components=n_cols, random_state=0)
pca.fit(X=X)
with np.printoptions(precision=3):
print(f'pca.components_=\n{pca.components_}')
print(f'pca.explained_variance_={pca.explained_variance_}')
print(f'pca.explained_variance_ratio_={pca.explained_variance_ratio_}')
print(f'pca.noise_variance_={pca.noise_variance_}')
print(f'pca.mean_={pca.mean_}')
print(f'pca.singular_values_={pca.singular_values_}')
```

Output:

```
pca.components_=
[[-0.219 -0.091 -0.752 -0.615]
  [ 0.854  0.439 -0.085 -0.265]
  [-0.41   0.882 -0.138  0.184]
  [-0.232  0.142  0.639 -0.72 ]]
pca.explained_variance_=[1.237 0.733 0.388 0.109]
pca.explained_variance_ratio_=[0.501 0.297 0.157 0.044]
pca.noise_variance_=0.0
pca.mean_=[-0.448  0.052 -0.093  0.247]
pca.singular_values_=[3.336 2.569 1.869 0.988]
```

What is the minimum number of principal components we must retain to account for 90% of the variability in the data? _____

(b) Suppose PCA is run on the data in the plot. Draw arrows on the plot representing the first two principal components. (There is more than one correct answer.)