This problem set has 10 questions, for a total of 100 points. Answer the questions below and mark your answers in the spaces provided. For all questions you **must show your work**, providing details on how your answer was calculated.

Your Name: _____

1. [5 points] For the following dataset, draw a Decision Tree of *minimum depth* that is consistent with the data.

x_1	x_2	y
0	0	No
0	1	No
1	0	Yes
1	1	Yes

2. Given the following dataset $\mathcal{D}$:

Midterm	Project	Attendance	y
High	Yes	High	Pass
Medium	Yes	High	Pass
Low	Yes	High	Pass
High	No	High	Pass
Low	Yes	Low	Fail
Medium	Yes	Low	Fail
Medium	No	Low	Fail
Low	No	Low	Fail

(a) [5 points] What is the *Entropy* of $\mathcal{D}$

(b) [10 points] Assuming the DT algorithm covered in lectures, what is the attribute picked as the root of the decision tree? Show the *information gain* of each attribute.

(c) [10 points] Assuming the DT algorithm covered in lectures, draw the final DT

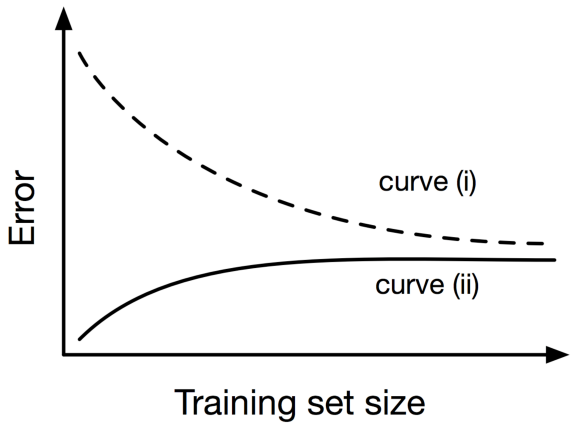
3. Let $\mathcal{X} = \{0,1\}^d$, $\mathcal{Y} = \{\mathrm{A},\mathrm{B},\mathrm{C},\mathrm{D},\mathrm{E}\}$, and $\mathcal{D}$ a dataset with 2000 instances equally distributed over all classes. Assume that the first feature is $\vec{x}_1 = 0$ for all instances labeled $\{\mathrm{A},\mathrm{B},\mathrm{C}\}$, and $\vec{x}_1 = 1$ for all instances labeled $\{\mathrm{D},\mathrm{E}\}$. You split the data into $\mathcal{D}_{train}$ (75%) and $\mathcal{D}_{test}$ (25%), preserving class distributions. Now consider the use of the following classifier:

$$f(\vec{x}) = \begin{cases} y & \text{if } \vec{x} \in \mathcal{D}_{train} \\ \mathrm{A} & \text{if } \vec{x} \notin \mathcal{D}_{train} \text{ and } \vec{x}_1 = 0 \\ \mathrm{D} & \text{if } \vec{x} \notin \mathcal{D}_{train} \text{ and } \vec{x}_1 = 1 \end{cases}$$

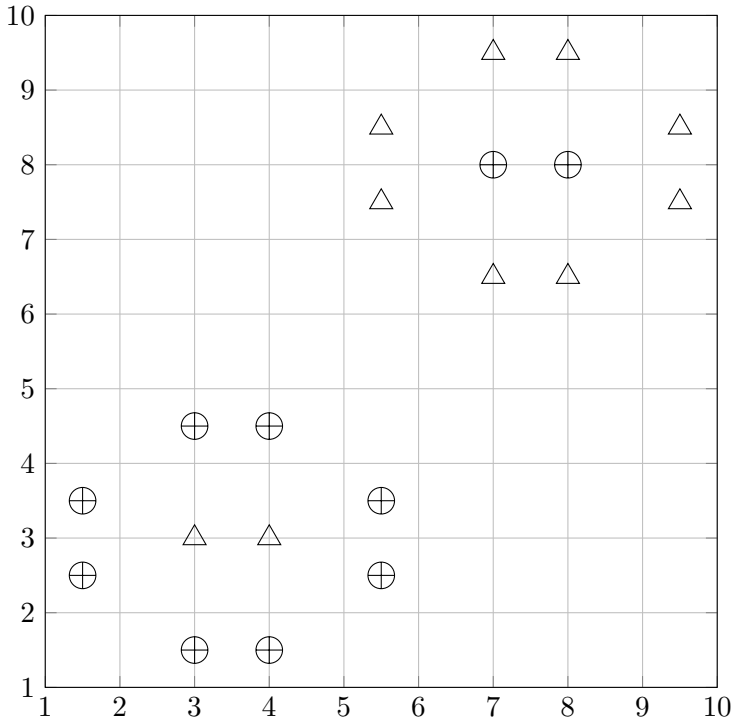
(a) [5 points] What is the 0/1 loss for the training set?

(b) [5 points] What is the 0/1 loss for the test set?

4. [5 points] The figure below shows a general trend of how the training and test errors change as we increase the training set size. Which curve best represents the training error? Justify your answer.



5. [10 points] Considering the dataset below and the use of *Euclidean distance*:



What value of k minimizes the Leave-One-Out Cross-Validation (LOOCV) error for a k-NN classifier?

6. [5 points] Mark each of the following as `T` if the statement is *true* and `F` if the statement is *false*

   (a) _____ Overfitting will be more likely when we reduce the size of the training set while keeping the same model complexity

   (b) _____ Overfitting is more likely for more complex hypothesis spaces

   (c) _____ Using k-fold cross-validation during training will guarantee the model does not overfit

   (d) _____ LOOCV generally gives more accurate estimates of the test error than 10-fold cross validation

   (e) _____ Cross-validation will guarantee that our model does not overfit

7. [5 points] Mark each of the following as `T` if the statement is *true* and `F` if the statement is *false*

   (a) _____ A 3-NN classifier is more robust to outliers than a 1-NN classifier

   (b) _____ *kd-trees* are used to reduce inference time by searching for approximate nearest neighbors

   (c) _____ Making a decision tree deeper will likely reduce training error and increase test error

   (d) _____ When pruning an already trained decision tree, we usually achieve better generalization

   (e) _____ If a decision tree performs badly on both training and test sets, it is possible that the tree is too shallow.

8. [5 points] Explain why using k-nn is not a good idea for very large datasets

9. Within the context of bias-variance decomposition:

   (a) [5 points] When you increase the number of neighbors k of a k-nn classifier, explain whether the bias will increase or decrease.

(b) [5 points] When you prune a decision tree, explain whether the variance will increase or decrease.

,

10. Assuming the following confusion matrix (rows are for predicted and columns for actual values):

	A	B
A	35	10
B	25	180

(a) [5 points] What is the *precision* of this model with respect to class A?

(a) _____

(b) [5 points] What is the *recall* of this model with respect to class A?

(b) _____

(c) [5 points] What is the overall *accuracy* of this model?

(c) _____

(d) [5 points] Alice can't tolerate A instances classified as B and Bob can't tolerate B instances classified as A. Using the precision and recall values you calculated before, which user will prefer this model more?