1. (a) Machine learning is a subset of artificial intelligence where models improve performance based on a task based on learning through data. Deep learning is a subset of machine learning that only utilizes a neural network with many layers.

   Artificial intelligence is an umbrella term to describe systems that display goal-directed, rational behavior similar to human intelligence. Machine learning is a method of representing the intelligence in AI and deep learning is a specialized method within machine learning to achieve the same goal.

   (b) i. The minimax algorithm is AI but not machine learning since the algorithm will choose the best available move that can be made based on some pre-defined utility function and doesn't require data or learning to do so.

   ii. k-nearest-neighbors (kNN) is machine learning but not deep learning because it has no hidden layers. Predictions come from distance comparisons to examples.

2. (a) Training and test splits estimate generalization of the model on unseen data. The model learns only from the training set while the test set is meant to be completely unseen. The model is evaluated on the test set after training to provide an unbiased estimate of performance on new data.

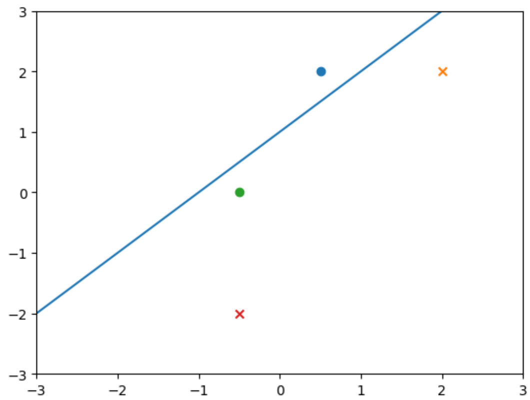
   (b) Overlap would leak information, inflating test accuracy. So, the training and test sets must be mutually exclusive.

   (c) Classification predicts a discrete category label, such as spam vs ham, and regression predicts continuous numeric values, such as the price of a house.

3. (a)

Step	Input (X_1, X_2)	Label y	Activation $a = \vec{w}^T\vec{x} + b$	Prediction	Weight/Bias Before	Weight/Bias After
0	-	-	-	-	$\vec{w} = [1, 1]$ $b = -1$	-
1	(1, 1)	-1	$1 + 1 - 1 = 1$	1	$\vec{w} = [1, 1]$ $b = -1$	$\vec{w} = [0.5, 0.5]$ $b = -1.5$
2	(-1, 1)	1	$-0.5 + 0.5 - 1.5 = -1.5$	-1	$\vec{w} = [0.5, 0.5]$ $b = -1.5$	$\vec{w} = [0, 1]$ $b = -1$
3	(-1, -1)	1	$0 - 1 - 1 = -2$	-1	$\vec{w} = [0, 1]$ $b = -1$	$\vec{w} = [-0.5, 0.5]$ $b = -0.5$
4	(1, -1)	-1	$0.5 - 0.5 - 0.5 = -0.5$	-1	$\vec{w} = [-0.5, 0.5]$ $b = -0.5$	no change

   (b)



Observation	(X_1, X_2)	True Y	$a = \vec{w}^T\vec{x} + b$	$\hat{Y} = sign(a)$
1	$(0.5, 2)$	1	$-0.25 + 1 - 0.5 = 0.25$	1
2	$(2, 2)$	-1	$-1 + 1 - 0.5 = -0.5$	-1
3	$(-0.5, 0)$	1	$0.25 + 0 - 0.5 = -0.25$	-1
4	$(-0.5, -2)$	-1	$0.25 - 1 - 0.5 = -1.25$	-1

(c) Confusion Matrix:

		Actual	
Predicted		1	-1
	1	1	0
	-1	1	2

$$
\begin{aligned}
Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} = \frac{1+2}{1+2+0+1} = 0.75 \\
Precision &= \frac{TP}{TP + FP} = \frac{1}{1+0} = 1 \\
Recall &= \frac{TP}{TP + FN} = \frac{1}{1+1} = 0.5
\end{aligned}
$$

(d) The confusion matrix shows balanced but mediocre results.

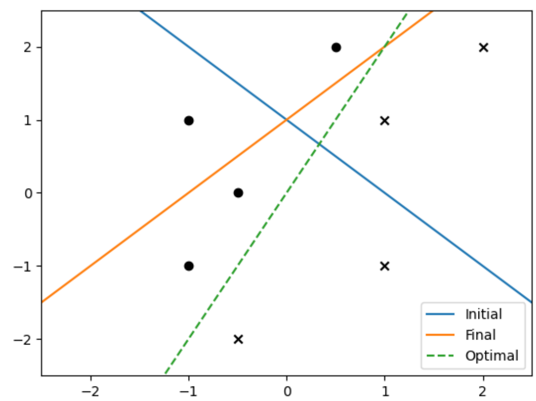
The perceptron is more accurate than not. An accuracy of 75% indicates some learning was done to achieve higher performance.

Precision of 100% indicates that the perceptron was fairly conservative in its predictions, having no errors in any positive predictions.

A recall of 50% shows the model misses half of all real positive. This risks importance instances being overlooked.

Together, the scores indicate some underfitting and room for improvement.

(e)



The second figure overlays three decision boundaries:

1. **Initial**: $X_2 = -X_1 + 1$
2. **Final**: $X_2 = X_1 + 1$
3. **Possible Optimal**: $X_2 = 2X_1$

Training rotated the decision boundary seemingly in a counter-clockwise direction so that the classes are more accurately classified. However, it seems like it stops short since it was approaching the possible optimal decision boundary, but still has some misclassified observations. A larger learning rate or more training cycles could have led this decision boundary to become nearly identical to the optimal decision boundary.