Exercise 1:

In contrast to the linear model(s) of traditional statistics, machine learning offers a broad range of alternative models to perform prediction on a data set. A problem where one wants to predict a categorical target variable given some feature variables is called **classification**.

The figure below plots three different classification models for the same data set. An observation with numerical features $\mathbf{x} = (x_1, x_2)^T$ can either belong to class 0 or class 1. The true class is indicated by the color of the corresponding point. The class a model predicts is indicated by the background color. For example, the single observation with $\mathbf{x} = (5, 0)^T$ belongs to class 0, which is correctly predicted by model A and model C, but incorrectly predicted by model B.
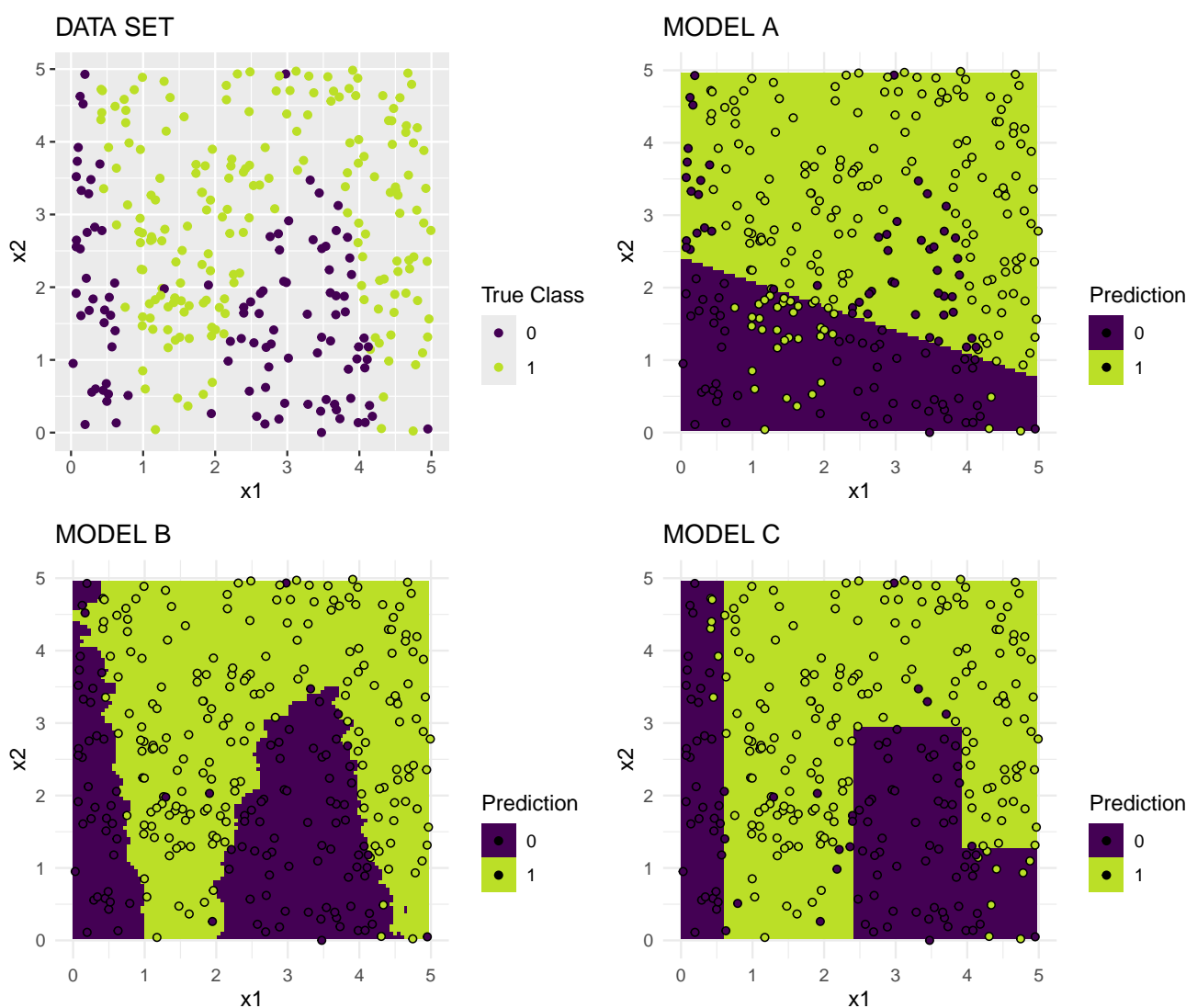
Each of the algorithms used to generate these **decision regions** (the set of input points where a certain class is predicted) will be discussed in detail in this course. For example, the decision region plotted for model A belongs to a logistic regression model trained on the data.

a) Assign the decision regions of model B and model C above to the two machine learning models described below. Try to explain your choice.

- **K-Nearest Neighbors (KNN) with k = 5**
  To predict the class of a new observation with features $\mathbf{x} = (x_1, x_2)^T$, find the 5 closest points to the new point in the data (by their euclidean distance) and select the class which is most common among these 5 points.

- **Classification and Regression Tree (CART)**
  To predict the class of a new observation with features $\mathbf{x} = (x_1, x_2)^T$, use the decision tree below, starting at the top. The colored circles at the bottom specify the class the model predicts.
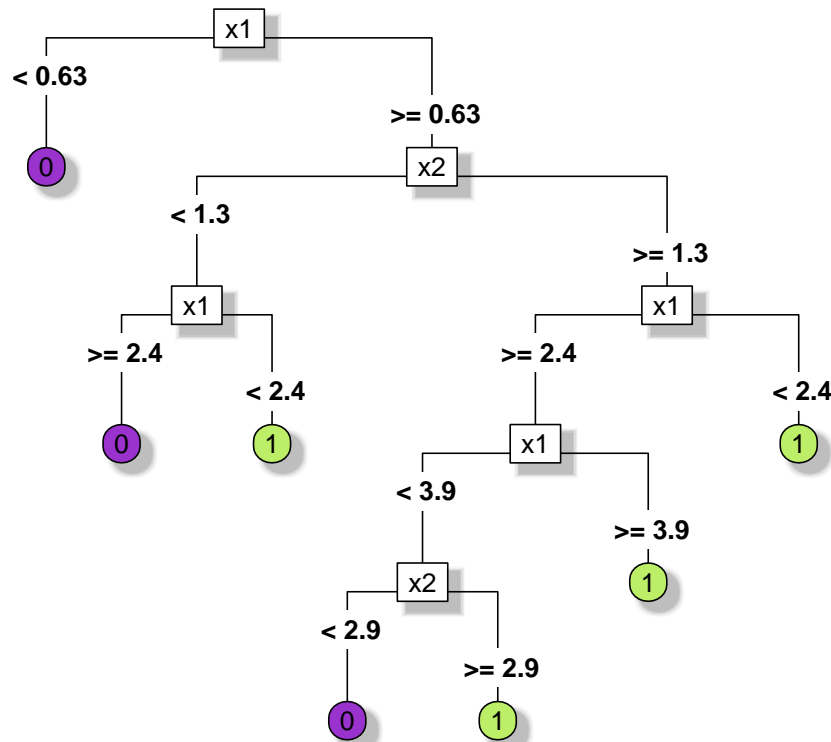
b) In machine learning, models can be evaluated in different ways. How could you compare the performance of the three models by using the confusion matrices below?

Model A

		Truth	
		1	0
Prediction	1	154	52
	0	37	57

Model B

		Truth	
		1	0
Prediction	1	185	9
	0	6	100

Model C

		Truth	
		1	0
Prediction	1	176	11
	0	15	98