Problem 1 (T, Decision Trees).

1. Sketch a tree corresponding to the partition of the predictor space indicated in Fig. 1. The numbers inside the boxes indicate the mean of Y within each region.

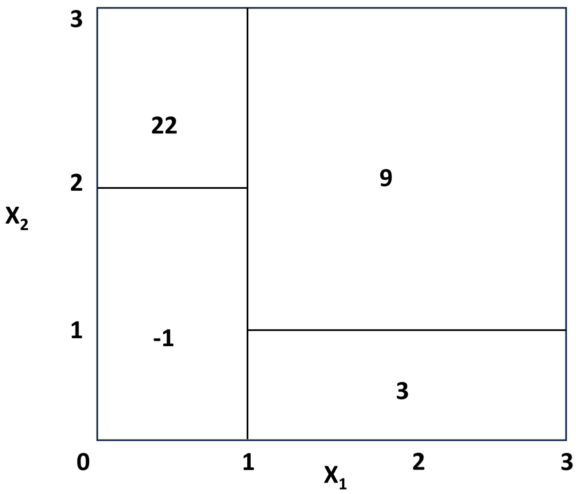

Figure 1: Predictor Space over variables X_1 and X_2

2. Now look at the tree shown in Fig 2. Sketch what the partition space for this tree would look like.

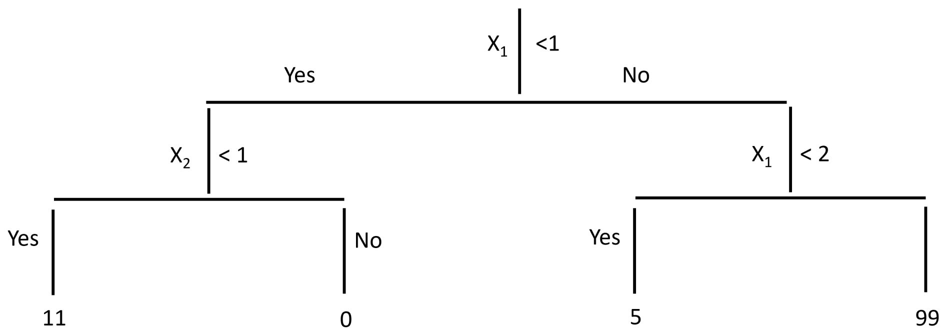

Figure 2: Decision tree over variables X_1 and X_2

Solution.

1. The predictor space results in the tree shown in Fig. 3.

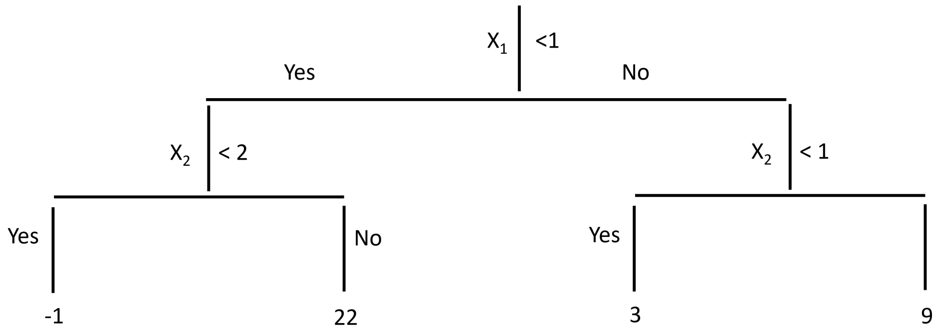

Figure 3: Decision tree over variables X_1 and X_2 for Part 2

2. The tree results in the predictor space shown in Fig. 4.

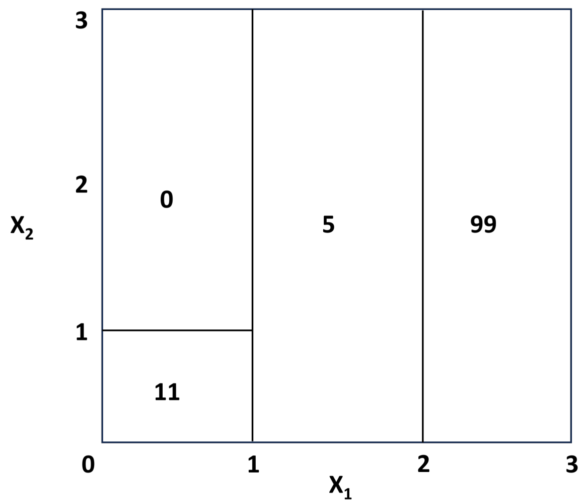

Figure 4: Predictor space variables X_1 and X_2 for Part 2

Problem 2 (T, Support Vector Machines).

1. We have seen that in $p = 2$ dimensions, a linear decision boundary takes the form $\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$. We now investigate a non-linear decision boundary.
   (a) Sketch the curve $X_1^2 - 2X_1 - X_2 = 0$.
   (b) On your sketch, indicate the set of points for which $X_1^2 - 2X_1 - X_2 > 0$ and the set of points for which $X_1^2 - 2X_1 - X_2 \leq 0$
   (c) Suppose that a classifier assigns an observation to the blue class if $X_1^2 - 2X_1 - X_2 > 0$, and to the red class otherwise. To what class are the following observations classified? $(0, 0), (-1, 1), (2, 2), (3, -8)$.
2. Consider the following optimization problem for Maximal Margin Classifier given in ISLP book Section 9.1.4. This classifier can be used to learn a linear decision boundary between two classes given that the classes are perfectly separable.

$$\text{maximize}_{\beta_0, \beta_1 \ldots, \beta_p, M} M \tag{2.1}$$

$$\text{subject to} \sum_{j=1}^{p} \beta_j^2 \; = \; 1, \tag{2.2}$$

$$y_i \left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} \right) \geq M \tag{2.3}$$

   (a) State whether the following statement is True or False: Constraint 2.2 is a constraint on the hyper-plane $(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} = 0)$.
   (b) Describe the purpose of Constraints 2.2 and 2.3.
3. In many cases no separating hyperplane exists, and so there is no maximal margin classifier. In this case, the optimization problem (2.1 - 2.3) has no solution with $M > 0$. Explain how you can extend the concept of a separating hyperplane to develop a hyperplane that *almost* separates the classes.

Solution.

1. (a) The sketch is shown in Fig. 5.

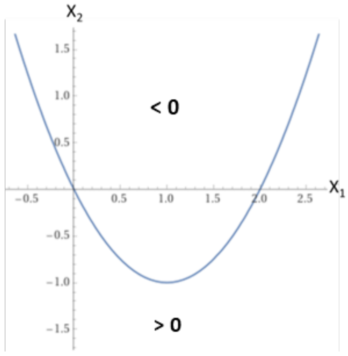

Figure 5: Sketch of $X_1^2 - 2X_1 - X_2 = 0$

   (b) All the points below the curve in Fig. 5 are ≥ 0, the boundary of the curve is $= 0$. Everything above the curve is < 0. Specified in Fig. 5

   (c) We plug in the values of X_1 and X_2 into the equations and compare the answer to the classification condition

   - $(0, 0) \rightarrow 0$. This is not greater than 0, hence this gets assigned to **Red**
   - $(-1, 1) \rightarrow 2$. This is greater than 0, hence this gets assigned to **Blue**
   - $(2, 2) \rightarrow -2$. This is not greater than 0, hence this gets assigned to **Red**
   - $(3, -8) \rightarrow 11$. This is greater than 0, hence this gets assigned to **Blue**

2. (a) False. If $\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} = 0$ defines a hyperplane, so does $k * (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} = 0)$ for a non-zero constant k. (Sec 9.1.4 ISLP).

   (b) Constraints 2.2 and 2.3 ensure that each observation is on the correct side of the hyperplane and at least a distance M from the hyperplane (Read Sec 9.1.4 ISLP).

   (c) By introducing slack variables ϵ_i and a maximum budget C to allow for severity of the violations to the margin (and to the hyperplane) that we will tolerate. (Sec 9.2 ISLP).