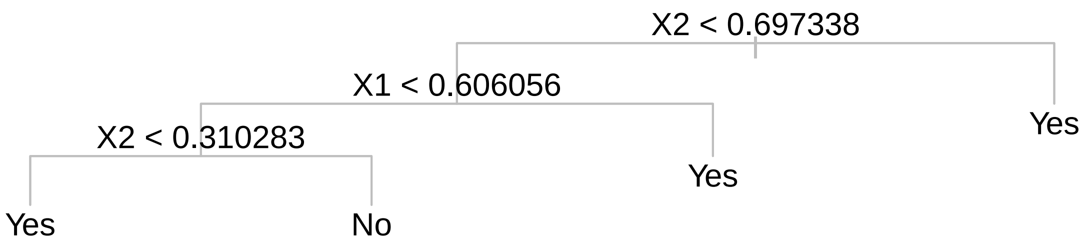
Introduction to Machine Learning

Second midterm test sample

The test

Q1 Classification

The classification tree was trained with response data Y and using two explanatory variables X_1, X_2.



A) Sketch the classification implied by the tree in the region $[0,1]^2$. Use "Yes"/1 and "No"/0.

B) A new set of points has been provided with fresh response values, as follows.

```
##        x1   x2 Y
## [1,] 0.25 0.33 1
## [2,] 0.50 0.33 0
## [3,] 0.75 0.33 0
## [4,] 0.25 0.67 0
## [5,] 0.50 0.67 0
## [6,] 0.75 0.67 1
```

Using the data and the trained tree, give predicted (0/1) labels for the new points.

C) Report the confusion matrix and performance measures TPR=☐ and FPR=☐. Briefly comment on the performance of this classifier.

D) Can the given trained tree be trimmed to a simpler version, removing redundancies?

E) Write a different tree that is equivalent to the given one.

Q2 Classification

A logistic classifier was fitted to data, and six fresh observations were used to assess the performance of this classifier. Here are the fitted logistic probabilities and the observations (Y values).

```
##      Logistic predictions Fresh data values
## [1,]                 0.90                 0
## [2,]                 0.27                 0
## [3,]                 0.37                 0
## [4,]                 0.57                 1
## [5,]                 0.93                 1
## [6,]                 0.20                 0
```

A) Compute the ROC curve for this classifier and report the coordinates of the curve. To this end, use thresholding to turn the predicted values into the 0/1 classes and for each threshold, compute confusion matrices and the usual performance measures (FPR,TPR). The list of all these performance measures gives the coordinates of the curve.

B) Compute the area under the curve and write it here AUC = $\boxed{}$. Briefly comment on the performance of this classifier.

Q3 Regression

Consider the lasso criterion as seen in the lectures $L = \frac{1}{2}Y^TY - \beta^TX^TY + \frac{1}{2}\beta^TX^TX\beta + \lambda||\beta||_1$. In this problem, you will consider the case with only one explanatory variable. The columns X and Y are centered so there is no intercept in the model.

The parameter β is a scalar and so are the quantities X^TX, X^TY and Y^TY. These scalars have values $X^TX = 2.5$, $X^TY = -0.5$ and $Y^TY = 2.5$.

A) Write the ordinary least squares estimate $\hat{\beta} = \boxed{}$.

B) Using the given quantities and the result of $\hat{\beta}$, write the lasso criterion as a function of β and λ (both scalars). The criterion is $L = \boxed{}$.

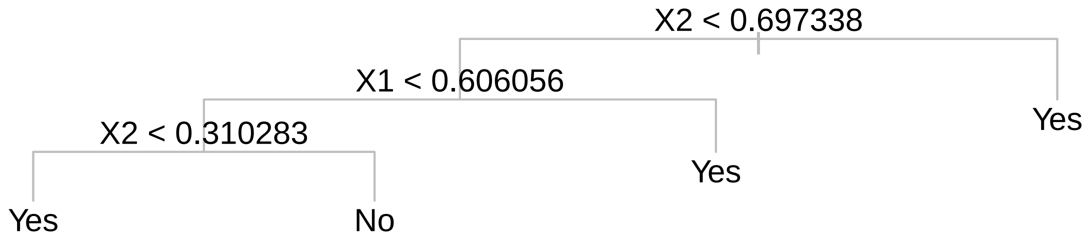
C) By doing derivative of L and solving the corresponding system, determine the lasso coefficient as a function of λ. That is $\hat{\beta}^L = \hat{\beta}^L(\lambda) = \boxed{}$.

D) The value of λ at which the path $\hat{\beta}^L(\lambda)$ shrinks to zero is $\lambda = \boxed{}$.

The solution

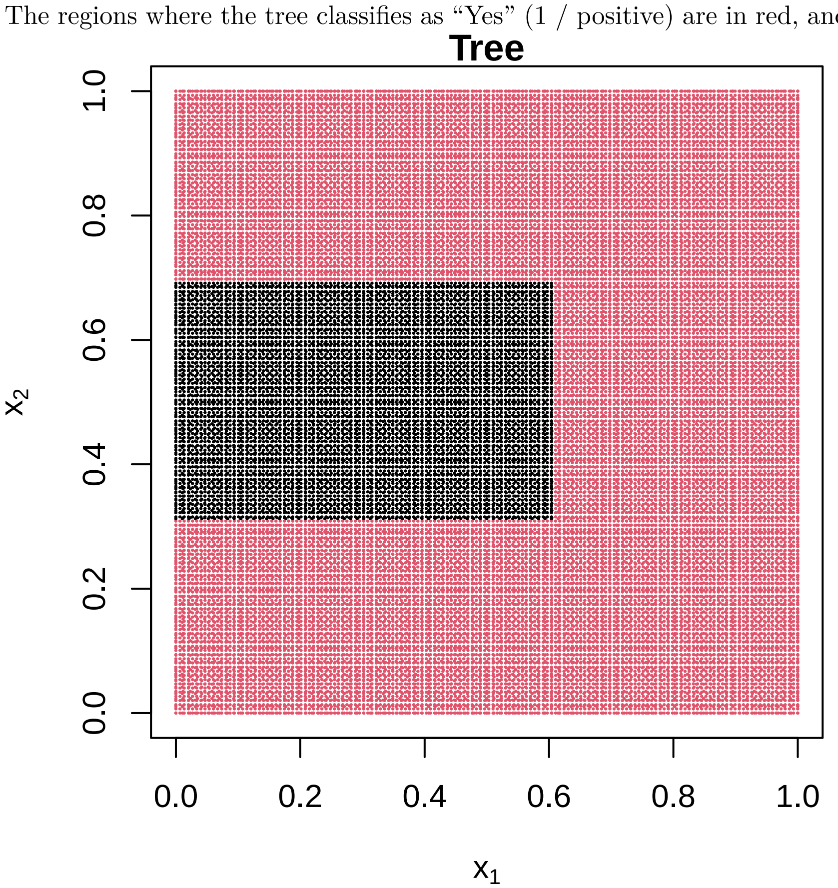
Q1 Classification

The classification tree was trained with response data Y and using two explanatory variables X_1, X_2.



A) Sketch the classification implied by the tree in the region $[0, 1]^2$. Use "Yes"/1 and "No"/0.

The regions where the tree classifies as "Yes" (1 / positive) are in red, and black color for "No" (0 / negative).



B) A new set of points has been provided with fresh response values, as follows.

```
##        x1   x2 Y
## [1,] 0.25 0.33 1
## [2,] 0.50 0.33 0
## [3,] 0.75 0.33 0
## [4,] 0.25 0.67 0
## [5,] 0.50 0.67 0
## [6,] 0.75 0.67 1
```

Using the data and the trained tree, give predicted (0/1) labels for the new points.

Predictions Ypr together with observed data are

```
##        x1   x2 Y Ypr
## [1,] 0.25 0.33 1   0
## [2,] 0.50 0.33 0   0
## [3,] 0.75 0.33 0   1
## [4,] 0.25 0.67 0   0
## [5,] 0.50 0.67 0   0
## [6,] 0.75 0.67 1   1
```

C) Report the confusion matrix and performance measures TPR= 0.5 and FPR= 0.25. Briefly comment on the performance of this classifier.

This is the confusion matrix

```
##      Ypredicted
## Ytrue No Yes
##   No   3   1
##   Yes  1   1
```

This is a moderately good classifier. It has a good performance when classifying negatives and it is less good for the identification of positives.

D) Can the given trained tree be trimmed to a simpler version, removing redundancies?

No, as no subtree has the same terminal labels.

E) Write a different tree that is equivalent to the given one.

(I will leave this question for you to do)

Q2 Classification

A logistic classifier was fitted to data, and six fresh observations were used to assess the performance of this classifier. Here are the fitted logistic probabilities and the observations (Y values).

```
##      Logistic predictions Fresh data values
## [1,]                 0.90                 0
## [2,]                 0.27                 0
## [3,]                 0.37                 0
## [4,]                 0.57                 1
## [5,]                 0.93                 1
## [6,]                 0.20                 0
```

A) Compute the ROC curve for this classifier and report the coordinates of the curve. To this end, use thresholding to turn the predicted values into the 0/1 classes and for each threshold, compute confusion matrices and the usual performance measures (FPR,TPR). The list of all these performance measures gives the coordinates of the curve.

Thresholding turns predicted probabilities into 0/1 values. If we sort these probability values, we have a good starting point for thresholds. The sorted probabilities are 0.2, 0.27, 0.37, 0.57, 0.9, 0.93; for thresholds, pick intermediate values.

To start with, any threshold smaller than 0.2 corresponds to a purely liberal classifier for which TPR=1,FPR=1.

Any value in the range (0.2, 0.27) can be used. If we take 0.24 as threshold, we have the confusion matrix:

```
##      ypred
## ytrue 0 1
##     0 1 3
##     1 0 2
```

with diagnostics TPR=1, FPR=0.75.

The procedure continues. Here we give the summary details.

For threshold 0.32, we have diagnostics TPR=1, FPR=0.5 from confusion matrix

```
##      ypred
## ytrue 0 1
##     0 2 2
##     1 0 2
```

For threshold 0.47, we have diagnostics TPR=1, FPR=0.25 from confusion matrix

```
##      ypred
## ytrue 0 1
##     0 3 1
##     1 0 2
```

For threshold 0.74, we have diagnostics TPR=0.5, FPR=0.25 from confusion matrix

```
##      ypred
## ytrue 0 1
##     0 3 1
##     1 1 1
```

For threshold 0.92, we have diagnostics TPR=0.5, FPR=0 from confusion matrix

```
##      ypred
## ytrue 0 1
##     0 4 0
##     1 1 1
```

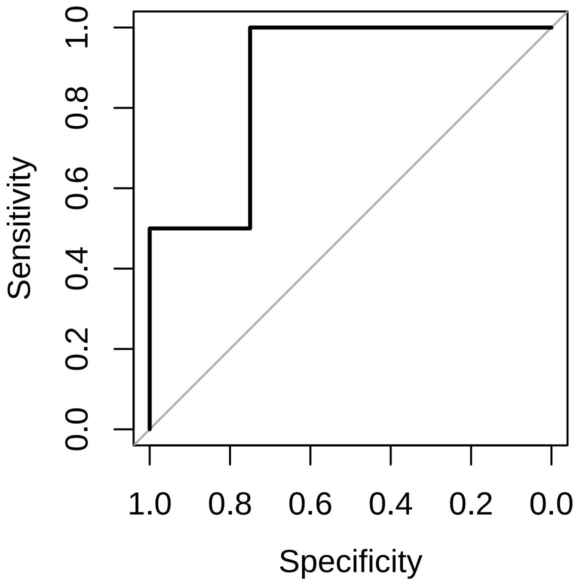
After the threshold exceeds 0.93, the classifier is a purely conservative with TPR=0, FPR=0 and the construction of the ROC curve ends. In summary, we have the following coordinates.

```
##      Sensitivity Specificity
## [1,]         1.0        0.00
## [2,]         1.0        0.25
## [3,]         1.0        0.50
## [4,]         1.0        0.75
## [5,]         0.5        0.75
## [6,]         0.5        1.00
## [7,]         0.0        1.00
```

equivalently

```
##      TPR  FPR
## [1,] 1.0 1.00
## [2,] 1.0 0.75
## [3,] 1.0 0.50
## [4,] 1.0 0.25
## [5,] 0.5 0.25
## [6,] 0.5 0.00
## [7,] 0.0 0.00
```

Here is the ROC (not asked but we can plot it).



B) Compute the area under the curve and write it here AUC = 0.875. Briefly comment on the performance of this classifier.

This is a good classifier which for small validation data gets quite close to the ideal classifier. If we were to suggest a point in the ROC curve, we would take (TPR,FPR)=(1, 0.25) as the best point from this classifier.

Q3 Regression

Consider the lasso criterion as seen in the lectures $L = \frac{1}{2}Y^TY - \beta^TX^TY + \frac{1}{2}\beta^TX^TX\beta + \lambda||\beta||_1$. In this problem, you will consider the case with only one explanatory variable. The columns X and Y are centered so there is no intercept in the model.

The parameter β is a scalar and so are the quantities X^TX, X^TY and Y^TY. These scalars have values $X^TX = 2.5$, $X^TY = -0.5$ and $Y^TY = 2.5$.

A) Write the ordinary least squares estimate $\hat{\beta} =$ -0.2.

B) Using the given quantities and the result of $\hat{\beta}$, write the lasso criterion as a function of β and λ (both scalars). The criterion is $L = 1.25 + 0.5\beta + 1.25\beta^2 - \lambda\beta$.

C) By doing derivative of L and solving the corresponding system, determine the lasso coefficient as a function of λ. That is $\hat{\beta}^L = \hat{\beta}^L(\lambda) = (\lambda - 0.5)/2.5 = 0.4\lambda - 0.2$.

D) The value of λ at which the path $\hat{\beta}^L(\lambda)$ shrinks to zero is $\lambda =$ 0.5.