Machine Learning

Author: Alec Peterson (ap3842@drexel.edu)

Assignment 6 - Probabilistic Models

Fall 2023

1 Theory

1. Consider the following set of training examples for an unknown target function: $(x_1, x_2) \rightarrow y$:

Y	x_1	x_2	Count
+	T	T	3
+	T	F	4
+	F	T	4
+	F	F	1
-	T	T	0
-	T	F	1
-	F	T	3
-	F	F	5

   (a) Compute the posteriors for the observation $x = [T, T]$ using:

      i. Inference (5pts)

      - $P(y = +|x_1 = T, x_2 = T) = \frac{P(y=+,x_1=T,x_2=T)}{P(x_1=T,x_2=T)} = \frac{3/21}{(3+0)/21} = \mathbf{1}$
      - $P(y = -|x_1 = T, x_2 = T) = \frac{P(y=-,x_1=T,x_2=T)}{P(x_1=T,x_2=T)} = \frac{0/21}{(3+0)/21} = \mathbf{0}$

      ii. Naive Bayes (5pts)

      - $\rho = [P(y = +) \cdot (P(x_1 = T|y = +) \cdot P(x_2 = T|y = +)]$
        $+[P(y = -) \cdot (P(x_1 = T|y = -) \cdot P(x_2 = T|y = -)]$

        $= [(\frac{12}{21}) \cdot (\frac{3+4}{12}) \cdot (\frac{3+4}{12})] + [(\frac{9}{21}) \cdot (\frac{0+1}{9}) \cdot (\frac{0+3}{9})] = 0.210$ (rounded)

      - $P(y = +|x_1 = T, x_2 = T) = \frac{P(y=+)\cdot(P(x_1=T|y=+)\cdot P(x_2=T|y=+))}{\rho}$

        $= \frac{(\frac{12}{21})\cdot(\frac{3+4}{12})\cdot(\frac{3+4}{12})}{0.210} = \mathbf{0.924}$

      - $P(y = -|x_1 = T, x_2 = T) = \frac{P(y=-)\cdot(P(x_1=T|y=-)\cdot P(x_2=T|y=-))}{\rho}$

        $= \frac{(\frac{9}{21})\cdot(\frac{0+1}{9})\cdot(\frac{0+3}{9})}{0.210} = \mathbf{0.075}$

2 Naive Bayes Classifier

Let's train and test a *Naive Bayes Classifier* to classifiy the fetal state from the Cartiotocography dataset.

In your report you will need:

1. Description of any additional pre-processing of the dataset you did.

   In training set, continuous columns in dataset were all made into binary categorical columns by comparing against respective mean of each column. If value in column was less than mean, made to 0, else made to 1.

   In validation set, continous columns were all similarly made into binary categorical columns by comparing against respective mean of each column from training set. If value in column was less than training mean, made to 0, else made to 1.

2. The validation accuracy of your system.

   Accuracy of validation set = 0.802

3. Your confusion matrix.

   $$\textbf{Confusion Matrix} = \begin{bmatrix} 450 & 70 & 35 \\ 11 & 73 & 7 \\ 3 & 14 & 45 \end{bmatrix}$$

3 Additional Dataset

(a) Description of any pre-processing of the observable data that you did.

No pre-processing performed on dataset besides required resizing of images to 40 x 40 pixels, then to 1 x 1600 arrays.

(b) Validation accuracy.

Accuracy of validation set = 0.804

(c) Confusion matrix.

Confusion Matrix for validation set =

$$\begin{bmatrix}
2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 3 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4
\end{bmatrix}$$