(1) (10 pts.) Give one major advantage (selling point) and one major disadvantage to using Neural Networks for predicting a binary target.

(2) (10 pts.) I have 5 features \( X_1, X_2, X_3, X_4, X_5 \) (centered and scaled) from which I compute the first principal component \( P_1 = c_1 X_1 + c_2 X_2 + c_3 X_3 + c_4 X_4 + c_5 X_5 \) by finding coefficients \( c \) that do what? In other words explain what finding \( P_1 \) accomplishes.

(3) (6 pts.) A logistic function can range from 0 to 1 making it useful for modeling probabilities. What is the range of the hyperbolic tangent function? ____ to _____. What node ___________________ in Enterprise Miner uses hyperbolic tangent functions?

(4) (15 pts.) I have a logistic model with logit \( L = 2 -5X_1 + 2X_2 \). For \( X_1 = 1 \) and \( X_2 = 3 \), compute the logit \( L = _____ \). What logit, \( L = _____ \), is associated with \( p = 1/2 \)? Keeping \( X_1 = 1 \), what new value of \( X_2 _____ \) would make \( p = 1/2 \)?

(5) I had 1000 people who made shampoo purchases in the hair care section of my store this year. Of these, 400 bought my store brand. Using the 1000 people as the root node of a probability tree (for estimating the probability of purchasing the store brand), I ran the tree analysis. The final tree had only 2 leaves, resulting from a single split on gender (no missing gender values) at the root node. In the male leaf, 160 people bought the store brand while 340 bought other brands.

(a) (6 pts.) Complete the resulting 2x2 contingency table.

<table>
<thead>
<tr>
<th></th>
<th>Store Brand</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>160</td>
<td>340</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) (5 pts.) I want to make a Bonferroni adjustment by comparing the p-value of the Chi-square statistic for this table to an adjusted version of the usual \( \alpha = 0.05 \) level. For this tree, I compute the adjusted \( \alpha \) value by dividing 0.05 by what number? ____________

(c) (6 pts.) Suppose, instead of a probability tree, I were developing a decision tree with this data. Would my two leaf tree change in some way now that my goal is to decide? (yes, no) Explain briefly.

(e) (15 pts.) If, based on my 2 leaf probability tree, I declare the 500 females to be buyers of the store brand but not the 500 males, what are the resulting sensitivity ______ and specificity ________? What is the cumulative lift _____ for my 2 leaf probability tree at the 50th percentile?
What are the distributional assumptions underlying Fisher’s Linear Discriminant analysis? Which, if any, of these assumptions is not needed when you use Fisher’s Quadratic Discriminant?

Suppose an ROC curve joins the (X,Y) points (0,0) (0.2,0.8) and (1,1) with line segments. Compute the area under the this ROC curve: C=____________

Why is this statistic used to assess your model’s ability to properly rank observations?

Besides the points (0,0) and (1,1) there is one more point that determines ROC chart for the 2 leaf tree. Compute that point’s horizontal _____ and vertical _______ coordinate.

1. Advantage: Can fit almost any surface.
   Disadvantage: Hard to understand and explain or tends to badly overfit on training data.

2. Choose c coefficients so that the single variable P captures as much variability in the entire set of features as possible.

3. H goes from -1 to 1 and is used in the Neural Network node (for the hidden units).

4. L=2-5(1)+2(3)=3, \( \logit=\ln\left(\frac{p}{1-p}\right)=\ln(1) = 0, \) \( 2-5+2X^2=0 \) implies \( X^2=1.5 \)

5. M 160 340 (note: prob. 160/500 = 0.32 < 0.50 for males)
   F 240 260 (note: prob. 240/500 = 0.48 < 0.50 for females and 0.32<0.48<0.50)

Only 1 choice for gender split so 0.05/1, (i.e. Bonferroni correction leaves it unchanged). Decision tree would not split. It would only have a root node because (see probabilities above) neither gender has more than 50% probability of purchasing the store brand.

Note that females are more likely to purchase the store brand than men. Based on our 2-leaf tree the possible rules are: say no one will purchase, say females will purchase but not males, say everyone will purchase. Recall that ROC compares what our decision rule says (500 buyers) versus what actually happened (400 buyers, some male, some female) and recall that gender is our only discriminating feature. We have to make the same decision for all females. We have no way to distinguish among the females as there were no further splits. Note also that the female set constitutes the most likely 50% which is why I need to say “ROC at the 50% level.” (do you see why the decision rule “males buy, females don’t” is not considered?)

For the stated decision rule (females buy, males don’t), the sensitivity is
   \( \Pr\{\text{say event | event }\} = \Pr\{\text{female | event }\} = \frac{240 \text{ females buyers}}{400 \text{ buyers}} = 0.60 \)
   (i.e. proportion of store brand buyers who are female)
Specificity = \Pr\{\text{say non event} \mid \text{non event}\} = \frac{\Pr\{\text{male} \mid \text{non event}\}}{600} = \frac{340}{600} = \frac{17}{30} = 0.5667. Since 50% are female and they have the higher probability of purchasing the store brand, the lift at 50% would be the probability of store brand purchase for females, 0.48, divided by the overall probability 0.40 for a lift of 1.2, (a 20% gain when you select the most likely 50%, that being the females, versus a random 50%)

The middle point on the ROC chart has horizontal coordinate 1 - specificity = 1 - 0.5667 = 1/3 and vertical coordinate = sensitivity = 0.60

6. Linear: Multivariate normal distributions, same variance-covariance matrix
   Quadratic: Variance-covariance matrices can differ.

7. Draw the ROC in a square with corners (0,0) and (1,1). The point (0.2, 0.8) is the upper left corner of a smaller square with area 0.8x0.8 = 0.64. The rest of the area consists of 2 triangles each, one to the left, one above the square, each with area 0.2x0.8/2 so we have area 0.08+.08+.64 = 0.80. Ranking is ordering. Concordance measures whether or not the model puts things in the right order. The area under the ROC curve is the proportion of concordant pairs plus half the proportion of ties so that is how it relates to concordance and hence to ranking.