BRAC University

CSE427: Machine Learning

Midterm Exam Summer 2023

Set - A, Duration: 1.5 hours

Answer any 3 of the following 4 questions.

1. **Problem 1: Linear Regression** [10 points] If a dataset contains m data instances and n features and a target column, if you want to do linear regression, then:

   (a) What should be the weights? [3 points]

   Solution: The weights are $w_1, w_2, \cdots, w_n, b$.

   (b) What should be the formula for output of the regression? [3 points]

   Solution: If x is a particular data instance, then the output for this instance is:

   $$y = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$$

   But for all the instances, the outputs should be:

   $$\begin{aligned} y^{(1)} &= w_1x_1^{(1)} + w_2x_2^{(1)} + \cdots + w_nx_n^{(1)} + b \\ y^{(2)} &= w_1x_1^{(2)} + w_2x_2^{(2)} + \cdots + w_nx_n^{(2)} + b \\ &\vdots \\ y^{(m)} &= w_1x_1^{(m)} + w_2x_2^{(m)} + \cdots + w_nx_n^{(m)} + b \end{aligned}$$

   Writing any of the two answers is fine. They can also be written in a matrix form.

   (c) Write the formula for the error function used in linear regression. [4 points]

   Solution: For a particular x, the error is:

   $$(y_{actual} - y_{predicted})^2$$

   For all the m instances, the total error should be:

   $$\sum_{i=1}^{m}(y_{actual}^{(i)} - y_{predicted}^{(i)})^2$$

   The following answers are also fine:

$\frac{1}{m}\sum_{i=1}^{m}(y_{actual}^{(i)} - y_{predicted}^{(i)})^2$

$\frac{1}{2}\sum_{i=1}^{m}(y_{actual}^{(i)} - y_{predicted}^{(i)})^2$

$\frac{1}{2m}\sum_{i=1}^{m}(y_{actual}^{(i)} - y_{predicted}^{(i)})^2$

$\frac{1}{2}(y_{actual} - y_{predicted})^2$

2. **Problem 2: Softmax Regression** [10 points]
Suppose, you have a dataset containing m data instances and each instance has n features. If x is a data instance, then the features are $x_1, x_2, \cdots, x_n$. There is another label column. The labels can be any integer from this list: $\{1, 2, 3, \cdots, k\}$. So, we are basically doing multi-class classification where there are k different classes.

(a) If we want to classify any data instance x, we have to calculate k different scores for k different classes. What are the formulae for that? [4 points]

Solution:

$$
\begin{aligned}
s_1 &= w_{1,1}x_1 + w_{2,1}x_2 + \cdots + w_{n,1}x_n + b_1 \\
s_2 &= w_{1,2}x_1 + w_{2,2}x_2 + \cdots + w_{n,2}x_n + b_2 \\
&\vdots \\
s_k &= w_{1,k}x_1 + w_{2,k}x_2 + \cdots + w_{n,k}x_n + b_k
\end{aligned}
$$

(b) Let's say that for data instance x, the scores calculated by our algorithm are $s_1, s_2, \cdots, s_k$. Now, how can we find the softmax probabilities for these scores? [4 points]

Solution:

$$
\begin{aligned}
p_1 &= \frac{e^{s_1}}{\sum_{i=1}^{k} e^{s_i}} \\
p_2 &= \frac{e^{s_2}}{\sum_{i=1}^{k} e^{s_i}} \\
&\vdots \\
p_k &= \frac{e^{s_k}}{\sum_{i=1}^{k} e^{s_i}}
\end{aligned}
$$

(c) If the true class of this data instance x is 3, then what is the loss incurred by this instance? [2 points]

Solution: We know that we use cross-entropy loss in soft-max regression and the formula is as follows:

$$-\sum_{i=1}^{k} y_i \log(p_i)$$

where $y = \{y_1, y_2, \cdots, y_k\}$ is a one-hot vector. Since the actual class is 3, only $y_3 = 1$ and the rest of them are 0. So, the above cross-entropy loss becomes:

$$-\log(p_3)$$

3. **Problem 3: Naive Bayes Classifier** [10 points]

(a) Suppose, we have 3 different discrete features and 1 label column in a dataset. The label can only be $TRUE$ or $FALSE$. For a new data $x = (x_1, x_2, x_3)$, write the formula for finding $P(label = TRUE|x = (x_1, x_2, x_3))$. [4 points]

Solution:
Writing any of the following two formulae is fine:

$$P(label = TRUE|x = (x_1, x_2, x_3)) = \frac{P(x=(x_1,x_2,x_3)|label=TRUE)P(label=TRUE)}{P(x=(x_1,x_2,x_3))}$$

$$P(label = TRUE|x = (x_1, x_2, x_3)) = \frac{P(x_1|label=TRUE)P(x_2|label=TRUE)P(x_3|label=TRUE)P(label=TRUE)}{P(x=(x_1,x_2,x_3))}$$

(b) What is the Naive Bayes assumption? [3 points]

Solution:
The naive-bayes classifier assumes that, all the features are conditionally independent with respect to the label. Hence, it uses the following equation:

$$P(x=(x_1,x_2,x_3)|label=TRUE) = P(x_1|label=TRUE)P(x_2|label=TRUE)P(x_3|label=TRUE)$$

(c) Let's say, feature x_1 can have only 3 different values and they are a, b, c. If the count of a is A, the count of b is B, the count of c is C, then what would be the Laplace smoothing probability for $P(x_1 = b)$? [3 points]

Solution:
If you use 1 for smoothing, then the answer is:

$$P(x_1 = b) = \frac{B+1}{A+B+C+3}$$

If you use any number k, then the answer is:

$$P(x_1 = b) = \frac{B+k}{A+B+C+3k}$$

4. **Problem 4: Decision Trees** [10 points]

(a) Let's say you invited 10 friends for a party in your house via Google calendar. 5 replied with COMING, 3 replied with NOT COMING, and 2 replied with UNSURE. What is the entropy here? [2 points]

Solution:

$$H = -\frac{5}{10}\log\frac{5}{10} - \frac{3}{10}\log\frac{3}{10} - \frac{2}{10}\log\frac{2}{10}$$

(b) Let's say you have a dataset and the entropy of the whole dataset is $H(S)$. You are splitting the dataset based on feature x_1. x_1 has only 3 values $\{a, b, c\}$ with counts A, B, C. When you split the dataset with $x_1 = a$, the entropy is H_a, when you split the dataset with $x_1 = b$, the entropy is H_b, when you split the dataset with $x_1 = c$, the entropy is H_c. What is the formula for information gain for features x_1? [4 points]

Solution:
When we want to find information gain, we need the initial entropy of the whole dataset which is $H(S)$. Then the question says, the dataset has been split based on feature x_1. x_1 has divided the dataset into three sets each having A, B, C data elements respectively. And each of these partitions have entropy H_a, H_b, H_c. So, the information gain should be:

$$InfoGain(x_1) = H(S) - \frac{A}{A+B+C}H_a - \frac{B}{A+B+C}H_b - \frac{C}{A+B+C}H_c$$

(c) What is the formula for the GainRatio of x_1 in this case? [4 points]

Solution:
We know that,

$$GainRatio(X) = \frac{InfoGain(X)}{SplitInfo(X)}$$

From the given prompt,

$$SplitInfo(x_1) = -\frac{A}{A+B+C}\log\frac{A}{A+B+C} - \frac{B}{A+B+C}\log\frac{B}{A+B+C} - \frac{C}{A+B+C}\log\frac{C}{A+B+C}$$

Therefore,

$$GainRatio(x_1) = \frac{H(S) - \frac{A}{A+B+C}H_a - \frac{B}{A+B+C}H_b - \frac{C}{A+B+C}H_c}{-\frac{A}{A+B+C}\log\frac{A}{A+B+C} - \frac{B}{A+B+C}\log\frac{B}{A+B+C} - \frac{C}{A+B+C}\log\frac{C}{A+B+C}}$$