ANSWER KEY

CS303L & CS519L Machine Learning - Test 2. Marks: 65. Instructor: Dr. Yeturu Kalidas
Date: 17/Oct/2023. Time: 16:15 - 17:15

QA) *[10 marks]* Fill in the blanks with correct answers for Binary Logistic Regression for output being {+1, -1} scenario. (Write all answers for this question in one page only)

There are N points in a dataset, $D = \{ (x_i, y_i) \}_{i \in [1, \dots, N]}$ where $y_i \in \{+1, -1\}$, $x_i \in R^M$. The dimensionality of input is __?(QA1)__. The dimensionality of the output is __?(QA2)__. The dimensionality of w vector is __?(QA3)__. The probability of prediction of $y = +1$ given w and x is written as __?(QA4)__. The expression $w \cdot x =$ __?(QA5)__. The probability of correct prediction for (x_i, y_i) is __?(QA6)__ as P(...) form. The probability of correct prediction in terms of exponentiation function is __?(QA7)__. The probability of correct prediction over the entire data set is equal to __?(QA8)__ (using π operator and P() function). Maximization of correct probability for (x_i, y_i) is the same as minimization of __?(QA9)__ (in terms of log function). The loss function for the data set is __?(QA10)__ (in terms of log function).

Ans.

A1) **M**	**A2)** **1**	**A3)** **M**	**A4)** $P(y = +1 \mid w, x)$	**A5)** $log(\frac{P(y=+1\mid w,x)}{P(y=-1\mid w,x)})$
A6) $P(y = y_i \mid w, x_i)$	**A7)** $\frac{1}{1+e^{-y_i w \cdot x}}$	**A8)** $\pi_{i=1}^{i=N} P(y = y_i \mid w, x_i)$	**A9)** $- \, log(P$ **(y=yi|w,xi))**	**A10)** $\sum_{i=1}^{i=N} log(1 + e^{-y_i w \cdot x})$

QB) Consider the two dimensional $y_i \in \{[1, 0]^T, [0, 1]^T\}$ corresponding to +1 and -1 case for binary logistic regression as in the above question, QA.

QB1) *[5 marks]* What is the corresponding predicted $\hat{y}_i$ (read as *yi hat*)?

QB2) *[5 marks]* Provide an expression for the *cross entropy* loss function for i-th data point. Use notation, $\hat{y}_i[0]$ and $\hat{y}_i[1]$ for 0th and 1st elements of the predicted output.

QB3) *[5 marks]* Explain the equivalence of loss functions constructed in QB2 and QA for a data point.

Ans.
QB1)

$\hat{y}_i = [\frac{e^{w \cdot x}}{1+e^{w \cdot x}}, \frac{1}{1+e^{w \cdot x}}]^T$

QB2)

$-\ (y_i[0] \times log(\hat{y}_i[0]) + y_i[1] \times log(\hat{y}_i[1]))$

QB3)

The loss function for output being vector, for case $y_i[0]$=1, $y_i[1]$=0 corresponds to yi = +1 case of $log(1\ +\ e^{-y_i(w \cdot x_i)})$

The same loss function for output being vector, for case $y_i[0]$=0, $y_i[1]$=1 corresponds to yi = -1 case of $log(1\ +\ e^{-y_i(w \cdot x_i)})$

QC) *[5 marks]* Given that $y_i\ \in \{0, 1\}$. How do you transform y_i to $t_i \in \{+\ 1, -\ 1\}$ formulation for binary logistic regression using a very simple transformation function?

Ans. $t_i\ =\ 2\ *\ y_i\ -\ 1$

QD) Consider a multi class logistic regression scenario, where $D\ =\ \{(x_i, y_i)\}_{i \in [1, \ldots N]}$ be N data points where $x_i \in R^A$ is the input and $y_i \in R^B$ is a one-hot encoded output corresponding to B classes. Let w_j denote the parameter vector for prediction of probability for j^{th} class. Let the pivot be the k^{th} class. Let $P(y\ =\ j|W, x)$ denote probability of prediction of output as j-th class where $W_{A \times B}$ is a matrix having $(\forall j\ \in\ [1,\ \ldots\ , B]\ )\ w_j\ =\ W[j,\ :\ ]$ the j-th row of the W matrix.

QD1) *[5 marks]* What is the expression for the probability of predicting j^{th} class in terms of the pivot class?

QD2) *[10 marks]* Derive the expression for the probability of predicting j^{th} class in terms of exponentiation operations? Neatly number the steps and follow perfect logical flow. Derivation should include about 4 or 5 steps.

Ans.

QD1)

$P(y\ =\ j|W, x)\ \ =\ \ e^{w_j \cdot x}\ \times\ P(y\ =\ k|W, x)$

QD2)

STEP 1: $\sum_{j=1}^{j=B} P(y = j|W, x) = 1$

STEP 2: $P(y = k|W, x) \times \sum_{j=1}^{j=B} e^{w_j \cdot x} = 1$

STEP 3: $P(y = k|W, x) = \frac{1}{\sum_{j=1}^{j=B} e^{w_j \cdot x}}$

STEP 4: $P(y = k|W, x) = \frac{1}{\sum_{j=1}^{j=B} e^{w_j \cdot x}}$

STEP 5: $P(y = j|W, x) = \frac{e^{w_j \cdot x}}{\sum_{m=1}^{B} e^{w_m \cdot x}}$

QF) Consider a binary classification scenario involving a data set having 7 points of which 4 are actual positives. Neatly illustrate for the following scenarios.

QF1) *[5 marks]* Give an example of predicted scores such that as the threshold increases the precision decreases.

QF2) *[5 marks]* In the same example you have constructed, as the threshold increases, how does the recall score behave?

Ans.

QF1) Precision values: 4/7, 3/6, 2/5, 1/4, 0/3, 0/2, 0/1

QF2) Recall values: 4/4, 3/ 4, 2/ 4, 1/ 4, 0/4, 0/4, 0/4

0.1	0.2	0.4	0.5	0.8	0.9	1.0
P	P	P	P	N	N	N

QG) *[10 marks]* Consider a data set having 5 points for the binary classification problem. Construct a simple setting of actual labels, non-zero probability scores, predicted labels and a non-zero threshold. Illustrate the following concepts of (i) Confusion matrix, (ii) TP, (iii) TN, (iv) FP, (v) FN, (vi) TPR, (vii) FPR, (viii) Recall, (ix) Precision and (x) Accuracy, all assuming non-zero values.

Ans. Any valid example is fine as we discussed in the lecture sessions.

ALL THE BEST