Name – Sahil Jiwan Ganbhoj

Branch – Electronics and Communication Engineering

Enrollment Number – BT20ECE022

Subject - Artificial Intelligence (Contlo Assignment)

Date – 17/12/23

Task 0

Contlo Assignment

Q/A Assignment

Q 1) Likely relationship between $W_{new0}, W_{new1}, \ldots, W_{newn}, W_{newn+1}$?

Ans: Case → Specific Model (Not be determined due to dependencies on other properties)
→ General Case

for Specific Model:-

- The relationship b/w weights will depend on factors like
  1) Regularization
  2) Correlation b/w features
  3) Model details
  4) Data set Used.

for General Case:- → Target Variable ~~Affect~~ Unaffected
→ Target Variable Affected.

If Target Variable is Unaffected:

- Duplicated feature has similar impact on Target Variable as by orizional.
- Therefore, we can expect rel"ship where W_{newn} & W_{newn+1} have similar values.

$$W_{newn} \approx W_{newn+1}$$

If Target Variable is Affected:

- Duplicate feature provides redundant information & is highly corelated with orizional feature, then

W_{new0} → Relatively Unchanged / Slight Change

W_{new1} → } Similar / Slightly Adjusted
W_{newn} →

W_{newn+1} → Decrease [To allocate share of predictive power b/w 2 duplicate features]

(Q.2) Which of the following is true?

Ans: On the first look, we can say option (A) is wrong. This is not true since sample size is constant for all templates & CTR of template (A) is 10%, which can be compared to CTRs of other templates.

For Calculating 95% Significance, we need Margin of Error.

$$\text{Margin of Error} = Z \times \sqrt{\frac{(SD)^2}{1000}}$$

where, $Z \rightarrow$ Critical value (1.96 for 95%)

$SD \rightarrow$ Standard Deviation (Assume $SD \cong 5\%$)

$$\therefore \text{Margin of Error} = 1.9 \times \sqrt{\frac{SD^2}{1000}} \cong 3.02\%$$

$\therefore$ Margin of Error $\cong 3\%$

- E's CTR (14%) is outside A's Margin of Error (10±3%), confirming it is significantly better.
- D's CTR (12%) falls ~~outside~~ inside A's Margin of Error, indicating its ~~superiority~~ ambiguous (Not different).

- B's CTR (7%) - falls out of A's Margin of error, showing its inferor to A by 95%.
- C's CTR (8.5%) needs longer testing to come to conclusion

Right Ans: → (B) E is better than A with 95% Confidence, B is worse than A with 95% Confidence. You need to run test for longer to tell where C & D compare to A with 95% Confidence.

Q.3) Computational cost of each gradiant descent iteration?

Ans:- Here, $m \rightarrow$ No. of training examples
$n \rightarrow$ No. of features
$k \rightarrow$ Avg. No. of Non-zero entries in each training example.

In logistic regression, computation cost depends on No. of Non-zero entries in features vectors.

• Evaluating gradient descent update for logistic regression.

① Computing dot product between feature vector & weight vector (n multiplication & additions)

② Apply sigmoid func^n (1 exponential func^n evaluation & 1 division)

③ Compute prediction error (1 subtraction)

$$\boxed{\text{Cost} = O((m \cdot k) + n)}$$

where, $m \cdot k \rightarrow$ Cost of evaluating dot product & applying sigmoid func^n for training ex.

$n \rightarrow$ Cost of computing prediction error & update weights for each feature.

Q.5) Find p using (a) MLE (b) Bayesian Estimate (c) MAP ?

Ans:- $P \rightarrow$ Probability of getting heads
$n \rightarrow$ No. of coin tosses
$k \rightarrow$ No. of heads.

(a) MLE of $P = \frac{k}{n} = \frac{7}{10} = 0.7$

(b) Bayesian Estimate of $P = \frac{k+1}{n+2} = \frac{7+1}{10+2} = \frac{8}{12} = 0.\overline{66}$
(α over $k+1$, β over $n+2$)

(c) MAP of P (Don't know!) (sorry!!!)
Hunch: MAP of $P = \frac{k+1}{n+2} = 0.67$

Q.4) Ranking models based on Accuracy.

Ans: Considering Pure Accuracy as criteria for evaluating these methods, likely ranking will be :-

① Method 3 : Pick random sample of 1 million stories from 1000 news sources & have them labelled. (Away from DB)
→ This method has most potential of increasing accuracy of system because it focuses on selecting stories where V1 classifier's o/p is both wrong & farthest away from decision boundary.
→ Best to handle edge cases & trains model to learn from mistakes & improve accuracy on wider range of data.

② Method 1 : Run V1 classifier on #1 million random stories from 1000 news sources. (O/p closest to decision boundary & get examples labelled).
→ Selecting ex. that are closest to decision boundary could provide some improvement but to a lesser extent.

③ Method 2 : Obtain 10k random labelled stories from 1k news
→ Randomly selecting training dataset may not ensure diverse range of challenging examples that can improve classifiers performance.

Based on the above analysis:-

Method 3 > Method 1 > Method 2

***