SuperAGI AI Assignment

Name: Shashank G (2022AIB2684)

December 1, 2023

1 Q/A Assignment

1. The equation for the first model in the dataset is:

$$P(y|x) = \sigma(\sum_{i=1}^{n} w_i x_i)$$

If there is a duplicated feature then the equation then becomes:

$$\begin{aligned} P(y|x) &= \sigma(w_0 x_0 + \cdots + w_n x_n + w_{n+1} x_{n+1}) \\ &= \sigma(w_0 x_0 + \cdots + (w_n + w_{n+1}) x_n) \end{aligned}$$

Therefore, weight w_n in the old model will be shared between w_n and w_{n+1} but the expressiveness of the model remains the same.

2. E is better than A with over 95% confidence, B is worse than A with over 95% confidence. You need to run the test for longer to tell where C and D compare to A with 95% confidence.

3. Generally when the data is not sparse the computation of the gradients is $O(n)$ for n features and since it is done for all m samples, the total complexity of the logistic regression is $O(n.m.num\_iters)$. But because of the sparse features only non-zero features will have gradient updates.

$$\Delta w_i = (y - \hat{y}) x_i$$

Since the gradient depends on the input features, the gradient will be zero if x_i is zero. Therefore we can keep note of only non zero features and update weight for those features only. The time complexity reduces to $O(k.m.num\_iters)$

4. While each method has its strengths and potential benefits, the actual performance of the V2 classifier will depend on the quality and representativeness of the labeled data. The manual labeling of challenging examples (Approaches 2 and 3) may offer better insights into areas where the V1 classifier struggles, potentially leading to improvements in accuracy for V2. However, the specific outcome can vary, and it's advisable to evaluate and compare the classifiers on a validation set to make informed decisions about their effectiveness.

5. (a) Log likelihood for parameter p considering all m samples is.

$$\ell(p) = \log \binom{n}{k} p^k (1-p)^{n-k}$$

Differentiating w.r.t to p and equating to zero, we get

$$\begin{aligned} \frac{k}{p} &= \frac{n-k}{1-p} \\ p &= \frac{k}{n} \end{aligned}$$

(b) Bayesian Estimate for data D is

$$\begin{aligned} P(p|D) &= \frac{P(D|p)P(p)}{P(D)} \\ &= \frac{\binom{n}{k}p^k(1-p)^{n-k}}{\int_0^1 \binom{n}{k}p^k(1-p)^{n-k}dp} \\ &= \frac{p^k(1-p)^{n-k}}{B(k+1, n-k+1)} \end{aligned}$$

This takes the form of a beta distribution whose mean is given $\frac{\alpha}{\alpha+\beta}$.
Therefore $p = \frac{k+1}{n+2}$

(c) MAP estimate: By differentiating the above p.d.f and equating it to 0, we get p which comes out to be k/n.

2 GPT code

In the following link: GPT code