Data Science Quiz 2

Instructor: Dr. Arian

Time: 90 minutes

---

1. Using Health Care Quality dataset, develop a logistic regression model and evaluate the results.

2. We consider the following models of logistic regression for a binary classification with a sigmoid function $g(z) = \frac{1}{1+e^{-z}}$:

   Model 1: $P(Y = 1 \mid X, w_1, w_2) = g(w_1 X_1 + w_2 X_2)$

   Model 2: $P(Y = 1 \mid X, w_1, w_2) = g(w_0 + w_1 X_1 + w_2 X_2)$

   We have three training examples:

   $$x^{(1)} = [1,1]^T \quad x^{(2)} = [1,0]^T \quad x^{(3)} = [0,0]^T$$
   $$y^{(1)} = 1 \qquad y^{(2)} = -1 \qquad y^{(3)} = 1$$

   a. Does it matter how the third example is labeled in Model 1 ? i.e., would the learned value of $\mathbf{w} = (w_1, w_2)$ be different if we change the label of the third example to -1 ? Does it matter in Model 2? Briefly explain your answer. (Hint: think of the decision boundary on 2D plane.)

   b. (Optional with extra scores) Now, suppose we train the logistic regression model (Model 2) based on the n training examples $x^{(1)}, \dots, x^{(n)}$ and labels $y^{(1)}, \dots, y^{(n)}$ by maximizing the penalized log-likelihood of the labels:

   $$\sum_i \log P\left(y^{(i)} \mid x^{(i)}, \mathbf{w}\right) - \frac{\lambda}{2} \| \mathbf{w} \|^2 = \sum_i \log g\left(y^{(i)} \mathbf{w}^T x^{(i)}\right) - \frac{\lambda}{2} \| \mathbf{w} \|^2$$

   For large λ (strong regularization), the log-likelihood terms will behave as linear functions of w.

$$\log\ g\left(y^{(i)}\mathbf{w}^T x^{(i)}\right) \approx \frac{1}{2} y^{(i)}\mathbf{w}^T x^{(i)}$$

Express the penalized log-likelihood using this approximation (with Model 1), and derive the expression for MLE $\hat{w}$ in terms of λ and training data $\{x^{(i)}, y^{(i)}\}$. Based on this, explain how w behaves as λ increases. (We assume each $x^{(i)} = \left(x_1^{(i)}, x_2^{(i)}\right)^T$ and $y^{(i)}$ is either 1 or -1)

In the following questions briefly explain the reason of your choice.

3. Is Logistic regression mainly used for Regression?

    a. TRUE

    b. FALSE

4. True-False: Is it possible to design a logistic regression algorithm using a Neural Network Algorithm?

    a. TRUE

    b. FALSE

5. Which of the following methods do we use to best fit the data in Logistic Regression?

    a. Least Square Error

    b. Maximum Likelihood

    c. Jaccard distance

    d. Both A and B