Optional

Subjective Questions - II

Q1. What is the Maximum Likelihood Estimator (MLE)?

The MLE chooses those sets of unknown parameters (estimator) that maximise the likelihood function. The method to find the MLE is to use calculus and setting the derivative of the logistic function with respect to an unknown parameter to zero, and solving it will give the MLE. For a binomial model, this will be easy, but for a logistic model, the calculations are complex. Computer programs are used for deriving MLE for logistic models.

(Here's another approach to answering the question.)

MLE is a statistical approach to estimate the parameters of a mathematical model. MLE and ordinary square estimation give the same results for linear regression if the dependent variable is assumed to be normally distributed. MLE does not assume anything about independent variables.

Q2. What are the different methods of MLE and when is each method preferred?

In the case of logistic regression, there are two approaches to MLE. They are conditional and unconditional methods. Conditional and unconditional methods are algorithms that use different likelihood functions. The unconditional formula employs the joint probability of positives (for example, churn) and negatives (for example, non-churn). The conditional formula is the ratio of the probability of observed data to the probability of all possible configurations.

The unconditional method is preferred if the number of parameters is lower compared to the number of instances. If the number of parameters is high compared to the number of instances, then conditional MLE is to be preferred. Statisticians suggest that conditional MLE is to be used when in doubt. Conditional MLE will always provide unbiased results.

Q3. What are the advantages and disadvantages of conditional and unconditional methods of MLE?

Conditional methods do not estimate unwanted parameters. Unconditional methods estimate the values of unwanted parameters also. Unconditional formulas can directly be developed with joint probabilities. This cannot be done with conditional probability. If the number of parameters is high relative to the number of instances, then the unconditional method will give biased results. Conditional results will be unbiased in such cases.

Q4. What is the output of a standard MLE program?

The output of a standard MLE program is as follows:

Maximised likelihood value: This is the numerical value obtained by replacing the unknown parameter values in the likelihood function with the MLE parameter estimator.

Estimated variance-covariance matrix: The diagonal of this matrix consists of the estimated variances of the ML estimates. The off-diagonal consists of the covariances of the pairs of the ML estimates.

Q5. Why can't we use Mean Square Error (MSE) as a cost function for logistic regression?

In logistic regression, we use the sigmoid function and perform a non-linear transformation to obtain the probabilities. Squaring this non-linear transformation will lead to non-convexity with local minimums. Finding the global minimum in such cases using gradient descent is not possible. Due to this reason, MSE is not suitable for logistic regression. Cross-entropy or log loss is used as a cost function for logistic regression. In the cost function for logistic regression, the confident wrong predictions are penalised heavily. The confident right predictions are rewarded less. By optimising this cost function, convergence is achieved.