Logistics Regression

1. What Are the Basic Assumption?
1. Linear Relation between independent features and the log odds

2. Advantages
1. Logistic Regression Are very easy to understand
2. It requires less training
3. Good accuracy for many simple data sets and it performs well when the dataset is linearly separable.
4. It makes no assumptions about distributions of classes in feature space.
5. Logistic regression is less inclined to over-fitting but it can overfit in high dimensional datasets.One may consider Regularization (L1 and L2) techniques to avoid over-fittingin these scenarios.
6. Logistic regression is easier to implement, interpret, and very efficient to train.

3. Disadvantages
1. Sometimes Lot of Feature Engineering Is required
2. If the independent features are correlated it may affect performance
3. It is often quite prone to noise and overfitting
4. If the number of observations is lesser than the number of features, Logistic Regression should not be used, otherwise, it may lead to overfitting.
5. Non-linear problems can't be solved with logistic regression because it has a linear decision surface. Linearly separable data is rarely found in real-world scenarios.
6. It is tough to obtain complex relationships using logistic regression. More powerful and compact algorithms such as Neural Networks can easily outperform this algorithm.
7. In Linear Regression independent and dependent variables are related linearly. But Logistic Regression needs that independent variables are linearly related to the log odds ($\log(p/(1-p))$).

4. Whether Feature Scaling is required?
yes

5. Missing Values
Sensitive to missing values

6. Impact of outliers?
Like linear regression, estimates of the logistic regression are sensitive to the unusual observations: outliers, high leverage, and influential observations. Numerical examples and analysis are presented to demonstrate the most recent outlier diagnostic methods using data sets from medical domain

Types of Problems it can solve(Supervised)
1. Classification

Practical Implementation
1. http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Performance Metrics
Classification
1. Confusion Matrix
2. Precision,Recall, F1 score