Revision Notes: Logistic Regression and Maximum Likelihood

Introduction to Logistic Regression

Logistic Regression is a statistical model used for binary classification problems where the output is a binary variable (i.e., true/false, yes/no, 0/1). Unlike linear regression, which predicts continuous outcomes, logistic regression predicts the probability of an outcome that can be defined in binary terms 【4:0†transcript.txt】.

Key Concepts:

- **Binary Classification**: Logistic regression is used when the outcome variable is binary. For example, predicting customer churn, where the result could either be 'Yes' (the customer leaves) or 'No' (the customer stays)【4:12†transcript.txt】.
- **The Logistic Function**: This is also known as the sigmoid function. It is represented mathematically as: $\sigma(z) = \frac{1}{1+e^{-z}}$ where *e* is Euler's number (approximately 2.718) 【4:3†transcript.txt】【4:16†transcript.txt】. The logistic function maps any real-valued number into the (0, 1) range, making it suitable for probability predictions.
- **Transformation of Variables**:
  - Input variables (X) are transformed into an intermediate variable (Z), often through a linear combination: $Z = W_1X_1 + W_2X_2 + ... + W_nX_n + W_0$ where Wifi's are the weights Interpreted in logistic regression【4:17†transcript.txt】【4:18†transcript.txt】.
  - Z is then input into the logistic function to calculate the probability that a given input belongs to a particular class 【4:8†transcript.txt】.

- The logistic function is smooth, continuous, and differentiable, making it suitable for optimization techniques such as gradient descent【4:4†transcript.txt】.
- The output of the logistic function can be treated as a probability, which is useful for classification decisions【4:19†transcript.txt】.

Gradient Descent and Maximum Likelihood Estimation (MLE)

- **Gradient Descent**: This is an optimization algorithm used to find the values of the parameters (*W*) that minimize a cost function. In the case of logistic regression, the cost function is related to the difference between the observed outcomes and the outcomes predicted by the model【4:16†transcript.txt】【4:7†transcript.txt】.
- **Maximum Likelihood Estimation**:
  - MLE is used to find the optimal parameters that make the observed data most probable. In logistic regression, this involves minimizing the log loss (a form of cost function)【4:5†transcript.txt】【4:6†transcript.txt】.
  - The objective is to determine the values of W that maximize the likelihood of the observed data【4:9†transcript.txt】【4:15†transcript.txt】.

Implementation and Use Cases

How Logistic Regression Works:

1. **Model the Data**: Start with a dataset with independent variables (features) and a dependent binary variable (target).
2. **Transform the Independent Variables**: Transform the input variables into the intermediate variable Z through a weighted sum【4:10†transcript.txt】.
3. **Apply the Logistic Function**: Convert Z into a probability using the logistic function.
4. **Predictive Decision**: Convert the probability into a binary outcome using a threshold (commonly 0.5)【4:17†transcript.txt】.

- When using logistic regression to predict customer churn, features might include total minutes a customer used the service, the number of calls made, etc. Based on these, the model predicts the probability of churn 【4:15†transcript.txt】.

Code Implementation:

- Logistic regression can be implemented using libraries such as sklearn in Python, which provides built-in functions for model fitting and prediction 【4:11†transcript.txt】【4:16†transcript.txt】.

Accuracy and Evaluation:

- **Accuracy**: It is calculated as the number of correct predictions divided by the total number of predictions 【4:6†transcript.txt】.
- Other metrics can also be used to assess the performance of a classification model, such as precision, recall, and the F1-score, which will be explored in future sessions 【4:16†transcript.txt】.

Overall, the class focused on understanding the theoretical and practical aspects of logistic regression, emphasizing its application in binary classification through transforming input features to obtain probabilities of class membership 【4:4†transcript.txt】 【4:13†transcript.txt】.