# BIG DATA CHAPTER 8 SUMMARY

**Author**
Kevin Loun

# Introduction

This document summarizes key concepts regarding the discriminative classifiers, focusing on logistic regression, and compares them with generative classifiers. It also details various optimization techniques used in fitting these models.

# Discriminative vs. Generative Classifiers

- **Generative classifiers** model the joint probability $p(y, x)$ and derive $p(y|x)$. Examples include Naive Bayes and Gaussian discriminant analysis.
- **Discriminative classifiers** model $p(y|x)$ directly, focusing only on the boundaries between classes. Examples include logistic regression and support vector machines.

# Logistic Regression

## Model Specification and Fitting

- Logistic regression models the probability of a binary outcome using a logistic function of linear predictors.
- Model fitting involves minimizing the negative log-likelihood or cross-entropy, using methods like gradient descent or Newton's method.

## Optimization Techniques

- **Gradient Descent**: Updates parameters in the direction of the steepest decrease in loss.
- **Newton's Method**: Utilizes the curvature of the loss function (second derivatives) to improve the convergence rates of parameter updates.
- **Quasi-Newton Methods**: Approximate the Hessian matrix to reduce computational costs, important in large-scale applications.

# Comparative Analysis

- Generative models are simpler to train and can naturally handle missing data and integrate unsupervised data.
- Discriminative models directly focus on the classification boundary, which can provide more accurate predictions and better calibrated probabilities.

## Conclusion

Both generative and discriminative models have their strengths and weaknesses. Choice of model depends on the specific requirements of the application, including the availability of data, the dimensionality of the input space, and the need for speed versus accuracy.