Aimee Co
MA189 SP24
Reading Summary 2

Summary of Murphy Chapter 8

Logistic regression models are a key method in machine learning for binary classification tasks. It distinguishes between generative and discriminative approaches to building probabilistic classifiers, with a focus on discriminative models that are linear in parameters for simplicity and efficiency in model fitting.

To estimate the parameters of a logistic regression model such as the non-closed form of the Maximum Likelihood Estimation (MLE), we need to use optimization algorithms to compute it. Thee gradient and Hessian of the negative log-likelihood, are crucial for optimization algorithms like gradient descent, Newton's method, and Iteratively Reweighted Least Squares (IRLS).

Gradient descent describes the basics of gradient descent optimization, including the challenge of choosing an appropriate step size. Newton's Method is a more advanced optimization technique that incorporates the curvature of the function being minimized, aiming for faster convergence by using both the gradient and the Hessian. IRLS is a specific adaptation of Newton's method for logistic regression, characterized by solving a weighted least squares problem at each iteration, highlighting its effectiveness in fitting logistic regression models. Quasi-Newton Methods can be used as alternatives to Newton's method, building up an approximation to the Hessian based on gradient information. L_2 regularization is important in preventing overfitting, especially in scenarios where data is linearly separable. This chapter also mentions the multi-class Logistic Regression and introduces the concept of softmax as a generalization of the logistic function for multiple classes.

Offline machine learning, online learning, and stochastic optimization are strategies for updating models in light of new data. Online Gradient Descent is an algorithm for updating model parameters step by step, using gradients and a step size, with possible projection to a constrained parameter space if necessary. Stochastic Optimization aims to minimize expected future loss, employing stochastic gradient descent (SGD) where updates are based on individual samples or mini-batches from the data stream. This chapter discusses some specific algorithms to achieve online learning:

1. LMS (Least Mean Squares): An online method for linear regression that adjusts parameters based on the error between predicted and actual values.
2. The Perceptron Algorithm: An early machine learning algorithm for binary classification, adjusting weights only on misclassified examples.
3. Bayesian Online Learning: Offers a probabilistic approach to online learning, updating the posterior distribution of parameters as new data arrives, allowing for a more nuanced understanding of model uncertainty and parameter variance.