BIG DATA CHAPTER 1 SUMMARY

Author
Kevin Loun

Introduction

Linear regression serves as a fundamental model in both statistics and supervised machine learning. It can adapt to non-linear relationships through basis function expansion and be used for classification by switching the Gaussian output to a Bernoulli or multinoulli distribution.

Model Specification

Linear regression models the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted x. The model is expressed as:

$$p(y|x, \theta) = N(y|w^T x, \sigma^2) \tag{1}$$

To model non-linear relationships, we can replace x with a non-linear function $\phi(x)$, keeping the model linear in the parameters w:

$$p(y|x, \theta) = N(y|w^T \phi(x), \sigma^2) \tag{2}$$

Maximum Likelihood Estimation (Least Squares)

The parameters of a linear regression model are typically estimated using maximum likelihood estimation (MLE), which is equivalent to minimizing the negative log-likelihood (NLL):

$$\theta_{\text{MLE}} = \arg\max_{\theta} \log p(D|\theta) \tag{3}$$

$$\text{NLL}(\theta) = -\sum_{i=1}^{N} \log p(y_i|x_i, \theta) \tag{4}$$

The NLL for Gaussian linear regression, where the negative log-likelihood is proportional to the residual sum of squares (RSS), is minimized:

$$\text{RSS}(w) = \sum_{i=1}^{N} (y_i - w^T x_i)^2 \tag{5}$$

Geometric Interpretation and Convexity

The least squares solution represents the projection of the response vector onto the space spanned by the predictors, which can be interpreted geometrically as minimizing the distance from the observed values to the predicted values in this space. The convex nature of the NLL ensures a unique global minimum, simplifying optimization.

Robust Linear Regression

To handle outliers, robust regression models like those using Laplace error terms (robust to outliers) or Huber loss function (combines squared and absolute errors) are used, allowing for errors that do not severely impact the regression fit:

$$p(y|x, w, b) = \text{Laplace}(y|w^T x, b) \propto \exp\left(-\frac{|y - w^T x|}{b}\right) \tag{6}$$