HW 3 Summary LEC 4: ML Algorithims

Lynnstacy Kegeshi

2025-01-31

Contents

1 Introduction

In this assignment, we summarize the subtopic on regression analysis in slide 4 of the Data Science Fundamentals notes. The course material covered regression models and how to interpret the results after fitting your regression model.

2 Regression Models and Residuals

Regression analysis is a statistical technique used to model and analyze relationships between variables. It helps in predicting a dependent variable Y based on one or more independent variables $(X_1, X_2, ..., X_p)$.

A regression model estimates a function that best fits the data. However, it is important to check if the model is correctly specified.

Parabolic Fit & Residuals - A parabolic (quadratic) fit means a second-degree polynomial regression was applied. - Residuals (differences between actual and predicted values) should appear random. If they do, the model is well-specified.

Adjusted R^2 vs R^2 - R^2 (coefficient of determination) measures how well the model explains variation in Y. - Adjusted R^2 (R^2_{adj}) is better when comparing models because it adjusts for the number of predictors, preventing overfitting.

2.1 Multiple Regression Model

The general multiple regression model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \epsilon$$

where:

- β_0: Intercept
- $\beta_1, \beta_2, ..., \beta_p$: Regression coefficients
- ϵ: Error term

Error variance estimation:

$$S^2 = \frac{RSS}{n - p - 1}$$

where:

- RSS = Residual Sum of Squares
- n = Sample size
- p = Number of predictors

R^2 and adjusted R^2:

$$R^2 = \frac{SS_{\text{reg}}}{TSS} = 1 - \frac{RSS}{TSS}$$

$$R^2_{\text{adj}} = 1 - \frac{RSS/(n - p - 1)}{TSS/(n - 1)}$$

2.2 Correlated Errors and ARX(1) Model

When errors in regression are correlated over time, this leads to autocorrelation.

Autoregressive model:

$$y_t = \alpha y_{t-1} + u_t$$

- Small samples (T) lead to biased estimates.
- When $|\alpha| > 1$, the model is not practical.

Regression Examples with Different α - $\alpha = -0.5$: Negatively correlated, useful in signal modeling. - $\alpha = 0.5$: Positively correlated, applied in seismological modeling. - $\alpha = 1$: Unit root process, common in stock market movements. - $\alpha = 1.2$: Explosive behavior, not useful in practice.

2.3 Exponential Regression

In exponential regression, the response variable Y changes at a rate proportional to itself:

$$\frac{dy}{dx} = ry$$

The solution to this differential equation is:

$$y = \alpha e^{rx}$$

- If $r > 0$, the function grows exponentially (e.g., population growth).
- If $r < 0$, it decays exponentially (e.g., radioactive decay).

Taking the natural log on both sides:

$$\log y = \log \alpha + rx$$

This makes it linear in log-space, allowing for linear regression.

2.4 Logistic Regression

Logistic regression is used when the response variable Y is categorical, usually binary (e.g., yes/no, success/failure).

Unlike exponential growth (which is unlimited), logistic regression accounts for a carrying capacity k, meaning that growth slows as it approaches a limit.

The equation for logistic regression is:

$$\log\left(\frac{y}{k-y}\right) = rx + c$$

The logistic function is:

$$\theta(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

This ensures that $\theta(x)$ remains between 0 and 1.

Logit Link Function

Taking the logit transformation:

$$\log\left(\frac{\theta(x)}{1-\theta(x)}\right) = \beta_0 + \beta_1 x$$

where:

- **Odds**: Ratio of success probability over failure probability.
- **S-shaped function**: Plot of $\theta(x)$ is sigmoid, but the logit plot is linear.

2.4.1 Maximum Likelihood Estimation (MLE)

In logistic regression, we estimate the coefficients β_0, β_1 using Maximum Likelihood Estimation (MLE). Given n independent observations:

$$P(Y_i = y_i | x_i) = \binom{y_i}{m_i} \theta(x_i)^{y_i} (1 - \theta(x_i))^{m_i - y_i}$$

Likelihood Function:

$$L = \prod_{i=1}^{n} \binom{y_i}{m_i} \theta(x_i)^{y_i} (1 - \theta(x_i))^{m_i - y_i}$$

Log-Likelihood Function:

Taking the logarithm of the likelihood function:

$$\log L = \sum_{i=1}^{n} [y_i \log \theta(x_i) + m_i \log(1 - \theta(x_i))]$$

We estimate β_0 and β_1 by maximizing this log-likelihood.