Coefficient of Determination (R^2)

Understanding R^2 and its cousins

Key Question:

How well does our model fit the data?

Overview:

- R^2 measures the proportion of variance explained
- It addresses the core question: "How good is my model?"

@AIinMinutes

Understanding Sums of Squares in Regression

Multiple Linear Regression Model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

where $\varepsilon_i \sim N(0, \sigma^2)$ and $\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ is our prediction.

Total Sum of Squares:

$$SS_{tot} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad \text{(Total variation in response)}$$

Residual Sum of Squares:

$$SS_{res} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad \text{(Unexplained variation)}$$

Regression Sum of Squares:

$$SS_{reg} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \quad \text{(Explained by predictors)}$$

1. Standard R^2

Proportion of variance explained by the model; measures goodness of fit for regression models

Formula:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Terms Explained:

- y_i: Observed values
- $\hat{y}_i$: Predicted values
- $\bar{y}$: Mean of observed values
- Range: 0 to 1 (higher is better)

Limitations of Standard R^2

Why the standard R^2 is not always sufficient? Doesn't account for overfitting.

Key Issues:

R^2 always increases (or stays the same) when adding predictors

Problems:

- Does not penalize model complexity
- Can lead to overfitting
- Makes more complex models appear better
- Does not address generalization to new data

@AIinMinutes

2. Adjusted R^2

Penalized version of R^2 accounting for model complexity; generally less than R^2

Formula:

$$\text{Adjusted } R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1}$$

Terms Explained:

- n: Sample size
- p: Number of predictors
- R^2: Standard coefficient of determination

@AIinMinutes

Likelihood Functions

Foundation for Generalized Linear model evaluation

Definition:

$$L(\theta|x) \propto P(x|\theta)$$

Log-Likelihood:

$$\ln(L) = \sum_{i=1}^{n} \ln(P(x_i|\theta))$$

Terms Explained:

- L: Likelihood function
- θ: Model parameters
- $P(x|\theta)$: Probability of observing data x given parameters
- Higher values indicate better fit

@AIinMinutes

The Normal Equation

Closed-form solution for multiple linear regression

Multiple Linear Regression Model:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Least Squares Objective:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 = \min_{\boldsymbol{\beta}} \|\mathbf{y} - X\boldsymbol{\beta}\|^2$$

Normal Equation Solution:

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{y}$$

Terms Explained:

- $\mathbf{y}$: Vector of response variables $(n \times 1)$
- X: Design matrix of predictors $(n \times p)$
- $\boldsymbol{\beta}$: Vector of coefficients $(p \times 1)$

@AIinMinutes

The Hat Matrix

Connecting predictions to the hat matrix

From Coefficients to Predictions:

$$\hat{\mathbf{y}} = X\hat{\beta} = X(X^TX)^{-1}X^T\mathbf{y}$$

The Hat Matrix:

$$H = X(X^TX)^{-1}X^T$$

Predicted Values via Hat Matrix:

$$\hat{\mathbf{y}} = H\mathbf{y}$$

Properties of H:

- Symmetric: $H^T = H$
- Idempotent: $H^2 = H$

@AIinMinutes

Leverage Points

Understanding influence in linear regression

Hat Matrix:

$$H = X(X^T X)^{-1} X^T$$

Leverage:

$$hii = [H]ii \text{ (diagonal elements of hat matrix)}$$

Terms Explained:

- X: Design matrix of predictors
- hii: Leverage (diagonal element of hat matrix); measures influence of observation i on predictions
- High leverage points strongly influence model fit

@AIinMinutes

3. Predictive R^2

Measures model's ability to predict for new observations using Leave-One Out CV

Formula:

$$\text{Predictive } R^2 = 1 - \frac{PRESS}{SS_{tot}}$$

PRESS Statistic:

$$PRESS = \sum_{i=1}^{n} (y_i - \hat{y}_{(i)})^2$$

Efficient Calculation:

$$y_i - \hat{y}_{(i)} = \frac{e_i}{1 - h_{ii}}$$

Terms Explained:

- $PRESS$: Prediction Error Sum of Squares
- $\hat{y}_{(i)}$: Prediction for observation i using model fitted without i
- e_i: Residual for observation i in the full model

4. Pseudo R^2

Alternatives for generalized linear models like logistic regression

Formula: For Nagelkerke's R^2:

$$R_N^2 = \frac{1 - \left(\frac{L_0}{L_M}\right)^{2/n}}{1 - L_0^{2/n}}$$

Terms Explained:

- L_M: Likelihood of the fitted model
- L_0: Likelihood of the null model
- n: Sample size
- Other variants: Cox & Snell, McFadden

@AIinMinutes