Logistic Regression: Reference Sheet

Purpose: Model the probability that a binary response $Y \in \{0,1\}$ occurs, based on one or more predictors $X_1, X_2, \ldots, X_k$, using the logistic function and Maximum Likelihood Estimation.

Model Formulation

Let $p(x) = P(Y = 1 \mid X = x)$. Logistic regression assumes:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k \quad \text{or} \quad p(x) = \frac{1}{1 + e^{-X^\top \beta}}$$

This is a Generalized Linear Model with:

- **Link function:** Logit (log-odds)
- **Mean function:** Logistic sigmoid $\sigma(z) = \frac{1}{1+e^{-z}}$

Interpretation of Parameters

- β_j: Change in log-odds of success per unit increase in X_j
- e^{β_j}: Odds Ratio (OR)
  - OR > 1: Positive effect
  - OR < 1: Negative effect
  - OR $= 1$: No effect

Deriving Parameters via Maximum Likelihood (Expanded)

Step 1: Likelihood Function

$$p_i = P(Y_i = 1 \mid x_i) = \frac{1}{1 + e^{-x_i^\top \beta}} \quad \Rightarrow \quad \mathcal{L}(\beta) = \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{1 - y_i}$$

Step 2: Log-Likelihood Function

$$\ell(\beta) = \sum_{i=1}^{n} \left[y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right]$$

Step 3: Gradient (Score Function)

$$\frac{\partial \ell}{\partial \beta} = X^\top (y - p)$$

Step 4: Hessian Matrix

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta^\top} = -X^\top W X \quad \text{where } W = \text{diag}(p_i(1-p_i))$$

Step 5: Newton-Raphson Algorithm (IRLS)

$$\beta^{(t+1)} = \beta^{(t)} - (X^\top W X)^{-1} X^\top (p - y)$$

Inference and Confidence Intervals

- **Standard Error (SE):** derived from $(X^\top W X)^{-1}$
- **Wald Test:** $z = \hat{\beta}_j / SE_j$
- **CI for β_j:** $\hat{\beta}_j \pm z^* \cdot SE_j$
- **CI for OR:** $\exp(\hat{\beta}_j \pm z^* \cdot SE_j)$
- **Likelihood Ratio Test:**
$$G^2 = -2\left[\log \mathcal{L}_{\text{reduced}} - \log \mathcal{L}_{\text{full}}\right] \sim \chi^2$$

Model Evaluation: Classification Metrics

Thresholding:

$$\hat{y}_i = \begin{cases} 1 & \text{if } \hat{p}_i \geq t \\ 0 & \text{otherwise} \end{cases} \quad \text{(commonly, } t = 0.5\text{)}$$

Confusion Matrix:

	Predicted 1	Predicted 0
Actual 1	TP	FN
Actual 0	FP	TN

Metrics:

- Accuracy $= \frac{TP+TN}{TP+FP+FN+TN}$
- Precision $= \frac{TP}{TP+FP}$
- Recall (Sensitivity) $= \frac{TP}{TP+FN}$
- Specificity $= \frac{TN}{TN+FP}$
- F1 Score = Harmonic mean of precision and recall

ROC Curve and AUC

ROC Curve: - Plots True Positive Rate (TPR) vs. False Positive Rate (FPR)

$$\text{TPR} = \frac{TP}{TP+FN}, \quad \text{FPR} = \frac{FP}{FP+TN}$$

AUC (Area Under Curve): - AUC = 0.5: Random - AUC = 1.0: Perfect

- 0.7–0.8: Acceptable
- 0.8–0.9: Excellent
- 0.9 +: Outstanding

Interpretation: AUC is the probability a randomly selected positive is ranked above a randomly selected negative.

Summary Table

Concept	**Interpretation**
β_j	Change in log-odds of success for one-unit increase in X_j
Odds Ratio (OR)	e^{β_j}; multiplicative change in odds
Standard Error (SE)	Estimated uncertainty in $\hat{\beta}_j$
Wald Test	Tests $H_0 : \beta_j = 0$ using z-score
Likelihood Ratio Test	Compare nested models using deviance
ROC Curve	Trade-off between TPR and FPR
AUC	Summary metric of classification performance