



Master of Science in Business Analytics

Machine Learning I
OPAN 6602

Week 3 Live Session

Instructor:
Tommy Jones, PhD

GEORGETOWN
UNIVERSITY / *MCDONOUGH*
SCHOOL of BUSINESS

“All models are wrong. Some Models are useful.”

– George Box

“An approximate answer to the right question is worth far more than a precise answer to the wrong one.”

– John Tukey

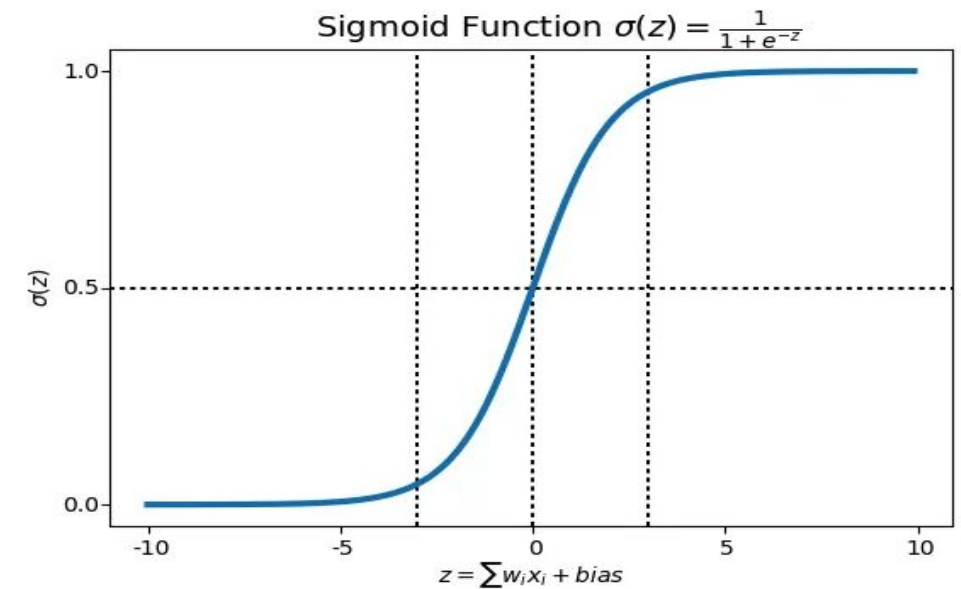
Bottom Line Up Front

1. Logistic regression is used for binary classification
2. Not as many diagnostics: rely on prediction to evaluate
3. Marginal effects:
exact odds and approximate for probabilities

Logistic Regression: Assumptions

1. Outcome is binary (two categories)
2. Predictor variables have a linear relationship to the log-odds of the outcome
3. Little or no multicollinearity of predictor variables

$$P(Y|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}$$



Sigmoid Function Graph

Evaluating Classifiers: Confusion Matrix

- **True Negative (TN):** Correctly predicted as negative
- **False Positive (FP):** Incorrectly predicted as positive
- **False negative (FN):** Incorrectly predicted as negative
- **True Positive (TP):** Correctly predicted as positive

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

- $Accuracy = \frac{TN+TP}{Total\ records}$

- $Sensitivity(Recall) = \frac{TP}{TP+FN}$

- $Precision = \frac{TP}{TP+FP}$

- **F-1 score:** harmonic mean of precision and recall

- $Specificity = \frac{TN}{TN+FP}$

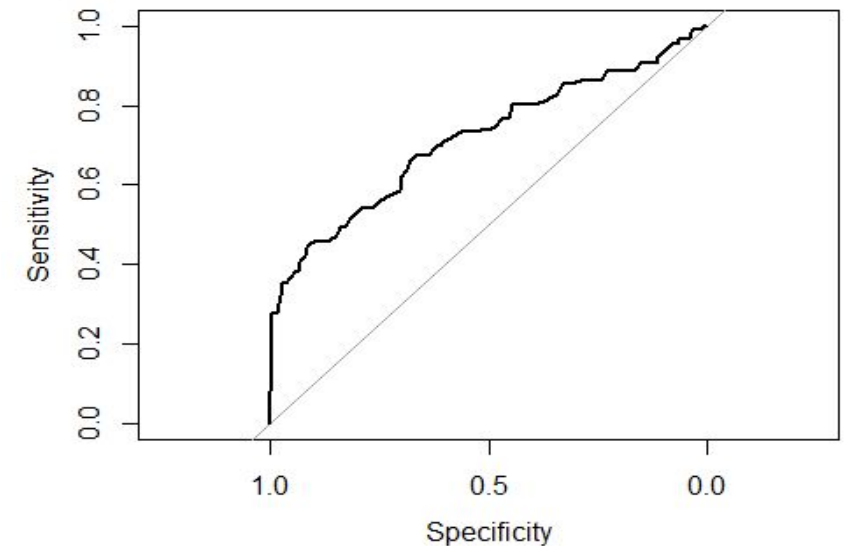
← Proportion of cases that are true that we correctly classified as true.

← Proportion of cases that we classified as true that are actually true.

← Proportion of cases that are false that we correctly classified as false.

Receiver Operator Characteristic Curve

- **Receiver Operating Characteristic Curve (ROC Curve)**
 - Plots Sensitivity (recall) against Specificity across a range of cutoff probabilities.
 - Area above the diagonal is better than random guessing.
- **Area under the ROC curve (AUC) is an overall performance measure.**
 - 0.5 = random guesses
 - 1 = perfect predictions
- **ROC curves are useful in comparing different models.**



Odds & Log Odds

Probability of $Y = 1$

$$P(Y|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}}$$

Odds ratio

$$\frac{P(Y|X)}{1 - P(Y|X)} = e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}$$

Log odds

$$\ln[P(Y|X)] - \ln[1 - P(Y|X)] = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

Marginal Effects 1

Remember: marginal effects are how we translate a model into plain language in the business context.

$$P(Y|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}$$

Exact marginal effect in odds:

$$\Delta\left[\frac{P(Y|X_k)}{1 - P(Y|X_k)}\right] = e^{(\beta_k)} - 1$$

Approximate. marginal effect in probability:

$$\Delta[P(Y|X_k)] \approx \frac{1}{N} \sum_{i=1}^N \frac{d\left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}\right)}{dx_{i,k}}$$

Marginal Effects 2

Remember: marginal effects are how we translate a model into plain language in the business context.

$$P(Y|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}$$

Exact marginal effect in odds:

“When $[X_k]$ increases by one unit, the odds of $[Y]$ increase/decrease by $[\#]$.”

Approximate. marginal effect in probability:

“When $[X_k]$ increases by one unit, the probability of $[Y]$ increases/decreases by $[\#]$ *on average*.”

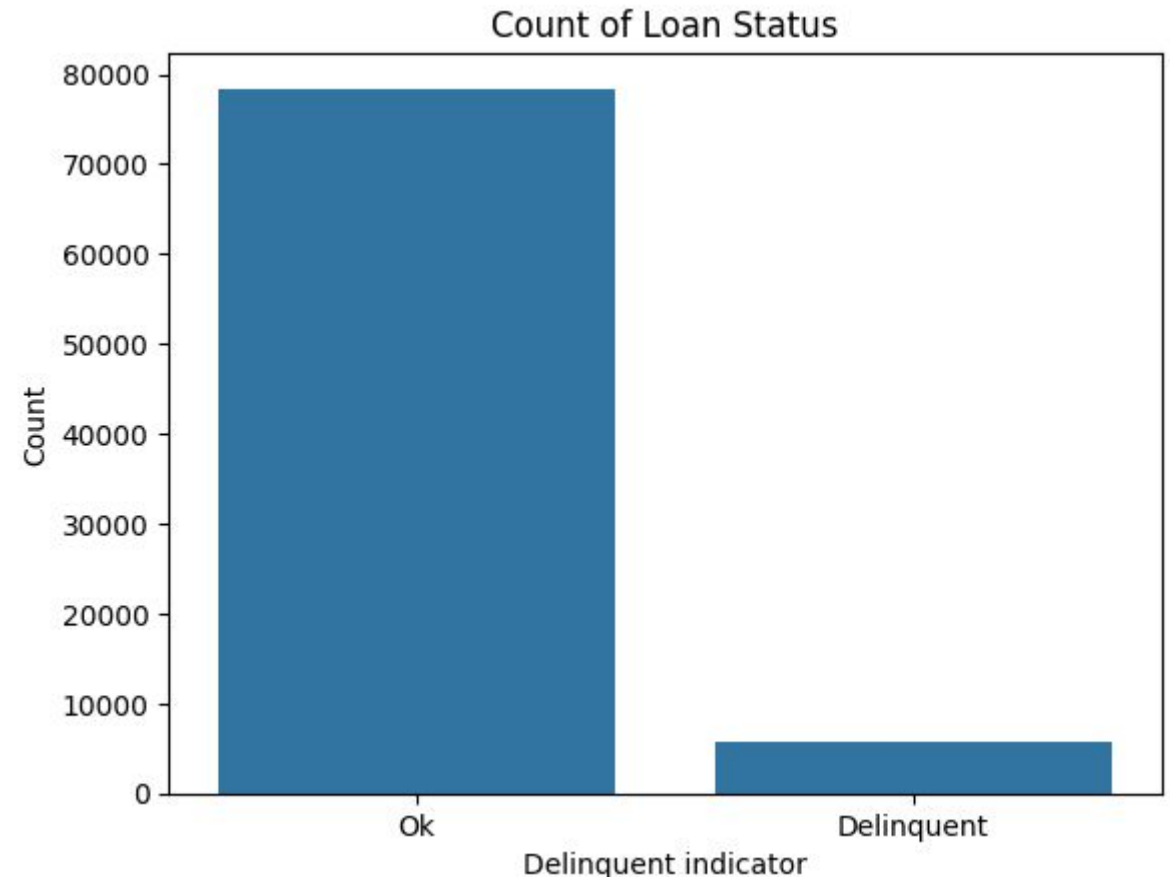
Practical Concern: Class Imbalances

Unequal representation: One or more classes appear far more often than others (e.g., 95% “no churn,” 5% “churn”).

Model bias: Standard algorithms minimize overall error, so they tend to predict the majority class and ignore the minority.

Misleading metrics: High accuracy can hide poor performance on the rare but important class — look at precision, recall, and F1 instead.

Balancing techniques: Re-sample the data (oversample minority / undersample majority) or re-weight the loss to give each class fair influence during training.



End