# Module 4: Classification

Lecture 9
Jan 30th, 2023

# Recap

- Classification
- Error for classification
- 
- KNN classifier
- Logistic Regression

# Recap

MICHIGAN STATE UNIVERSITY

- Classification
- Error for classification
- Bayes classifier

$$\text{Error}_{\text{Tr}} = \frac{1}{N} \sum_{i \in \text{Tr}} I[y_i \neq \hat{C}(x_i)]$$

$$C(x) = j \text{ if } p_j(x) = \max\{p_1(x), p_2(x), \ldots, p_K(x)\}$$

- KNN classifier
- Logistic Regression (Y can be either 0 or 1)

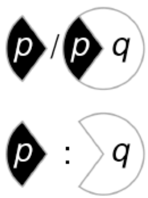
# Odds

$$\frac{p(x)}{1-p(x)} = \frac{\Pr(Y=1 \mid X=x)}{1-\Pr(Y=1 \mid X=x)} = \frac{\Pr(Y=1 \mid X=x)}{\Pr(Y=0 \mid X=x)}$$

- Logistic function is chosen so that odds are linear
- Can take any value from 0 (low odds) to $\infty$ (high odds)

Probability or risk $= \dfrac{p}{p+q}$

Odds $= p : q$



# Odds

- If the probability of default is 90%, what is the odds?

- If the odds is 1/3, what is the probability of default?

# Logistic Regression

MICHIGAN STATE UNIVERSITY

- Let $p(X) = \Pr(Y = 1|X)$ . How can we turn this to something range of a real line?

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X.$$

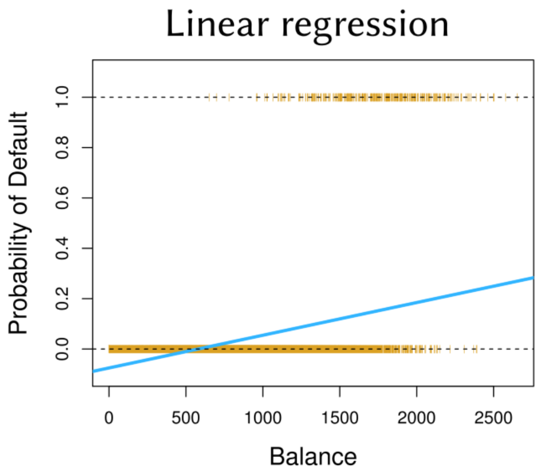
- This monotone transformation is called the log odds or logit

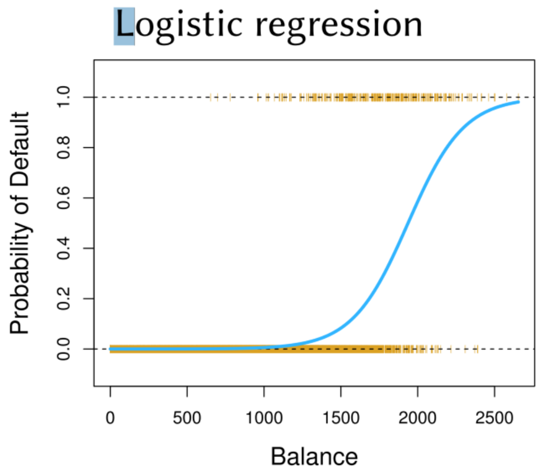
- Logistic regression uses the form

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}.$$

# Logistic Regression

$$\mathbb{P}[\text{default} = \text{yes} \mid \text{balance}] = \frac{e^{\beta_0+\beta_1 \text{balance}}}{1 + e^{\beta_0+\beta_1 \text{balance}}}$$



# Using Coefficient to predict

MICHIGAN STATE UNIVERSITY

| | Coefficient | Std. error | $z$-statistic | $p$-value |
|---|---|---|---|---|
| Intercept | $-10.6513$ | 0.3612 | $-29.5$ | $<0.0001$ |
| balance | 0.0055 | 0.0002 | 24.9 | $<0.0001$ |

What is the estimated probability of default for someone with a balance of $1,000?

$$\hat{p}(X) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 X}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 X}} = \frac{e^{-10.6513+0.0055\times 1000}}{1+e^{-10.6513+0.0055\times 1000}} = 0.006$$

What is the estimated probability of default for someone with a balance of $2,000:

$$\hat{p}(X) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 X}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 X}} = \frac{e^{-10.6513+0.0055\times 2000}}{1+e^{-10.6513+0.0055\times 2000}} = 0.586$$

# Interpreting the Coefficient

MICHIGAN STATE UNIVERSITY

| | Coefficient | Std. error | $z$-statistic | $p$-value |
|---|---|---|---|---|
| Intercept | −10.6513 | 0.3612 | −29.5 | <0.0001 |
| balance | 0.0055 | 0.0002 | 24.9 | <0.0001 |

What is the estimated probability of default for someone with a balance of $1,000?

$$\hat{p}(X) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 X}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 X}} = \frac{e^{-10.6513+0.0055\times 1000}}{1+e^{-10.6513+0.0055\times 1000}} = 0.006$$

What is the estimated probability of default for someone with a balance of $2,000:

$$\hat{p}(X) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 X}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 X}} = \frac{e^{-10.6513+0.0055\times 2000}}{1+e^{-10.6513+0.0055\times 2000}} = 0.586$$