Section 4D. Maximum Likelihood

Statistics for Data Science

Victor M. Preciado, PhD MIT EECS
Dept of Electrical & Systems Engineering
University of Pennsylvania
preciado@seas.upenn.edu

Maximum Likelihood Estimation (MLE)

Given a training dataset $\mathcal{D}_{\mathrm{Tr}} = \{(x_i, y_i)\}_{i=1}^{N}$ of independent samples with $y_i \in \{0,1\}$, we would like to estimate the parameters β_0 and β_1 in the logistic function that best explain the data

- **Maximum likelihood criterion**: We can use a criterion, called *maximum likelihood estimation* (MLE), to estimate the values of the parameters β_0 and β_1 from $\mathcal{D}_{\mathrm{Tr}}$
- Assume that the input variables $\{x_i\}_{i=1}^{n}$ are given. The *likelihood function* is defined as the conditional probability of observing a set of outputs $\{y_i\}_{i=1}^{n}$ given the inputs $\{x_i\}_{i=1}^{n}$, i.e,

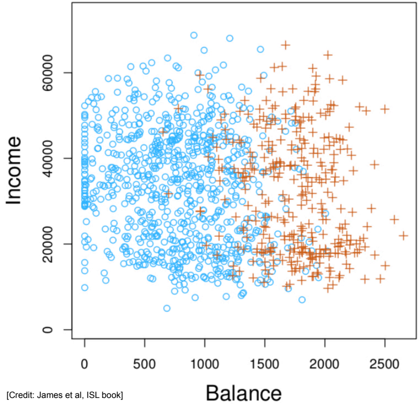
$$\begin{aligned}\ell(\theta_0, \theta_1) &= \prod_{i=1}^{N} \Pr(Y = y_i | X = x_i; \theta_0, \theta_1) \\ &= \prod_{i:\ y_i=1} p_1(x_i; \theta_0, \theta_1) \prod_{i:\ y_i=0} p_0(x_i; \theta_0, \theta_1)\end{aligned}$$

- We can estimate the unknown parameters using the MLE criterion:

$$\left(\widehat{\beta}_0, \widehat{\beta}_1\right) = \arg\max_{\theta_0, \theta_1} \ell(\theta_0, \theta_1)$$

MLE: Numerical Example

Consider the Default dataset in the figure below. Using Balance as the only input variable find the coefficients of a univariate logistic curve (as well as their standard deviations)



[Credit: James et al, ISL book]

[Credit: James et al, ISL book]	Coefficient	Std. Error	Z-statistic
Intercept	-10.6513	0.3612	-29.5
balance	0.0055	0.0002	24.9

Maximum Likelihood: Making Predictions

- Once we find $\widehat{\beta}_0$ and $\widehat{\beta}_1$, we can make predictions about the probability of an individual being in the **Default** class when we know her **Balance**:

$$\widehat{p}_1\left(\text{Balance}\right) = p_1\left(\text{Balance}; \widehat{\beta}_0, \widehat{\beta}_1\right) = \frac{e^{-10.65+0.0055\times\text{Balance}}}{1+e^{-10.65+0.0055\times\text{Balance}}}$$

  - For example, an individual with a Balance=\$1000 has a probability of defaulting of 0.006 (0.6%)
  - For example, an individual with a Balance=\$2000 has a probability of defaulting of 0.586 (58.6%)