
Classification

Suppose we want to build a mathematical relationship between:

- ▶ $y_i \in \{0,1\}$, denoting a player as a batsman or not a batsman.
- ▶ $X = (x_1, \dots, x_p) \in \mathbb{R}^{n \times p}$, which is a data matrix, with row i denoting a cricket player and column j denoting a variable, e.g. batting average

This is called a **classification problem** and not at all trivial!

Let's Take A Step Back First

- ▶ What if we want to predict the **highest number of runs** by **batting average**?

