




データと特徴量

データがテーブルに含まれている場合、各行はデータ点である

たとえば、動物のデータセットがある場合、各行は異なる動物を表している

ref: なっとく！機械学習 p13、p19

このテーブル内の各動物は、その動物の**特徴量**によって説明される

 **特徴量** モデルが予測を行うために使うことができるデータの特性や属性

データがテーブルに含まれている場合、特徴量はテーブルの列であり、特徴量は各データを説明する



予測とラベル

特徴量の中には、**ラベル**と呼ばれる特別なものがある

一般に、特定の特徴量を他の特徴量に基づいて予測しようとしているなら、その特徴量はラベルである

ref: なっとく！機械学習 p19~20

機械学習モデルの目標は、

データに含まれている**ラベル**を推測すること

であり、モデルが行う推測を**予測**と呼ぶ



ラベル付きデータとラベルなしデータ

データには、大きく分けて、

ref: なっとく！機械学習 p20～21

- **ラベル付きデータ** : ラベルが付いているデータ
- **ラベルなしデータ** : ラベルが付いていないデータ

の 2 種類がある

予測したいと思うような列を持たないデータセットは、ラベルなしデータである

ラベル付きデータとラベルなしデータは、**教師あり学習**と**教師なし学習**という 2 種類の機械学習を生み出している