機械学習プロセス
徹底解説

Step-5　Idea-1

16t4032n　田中裕隆

STEP－５　特徴エンジニアリング

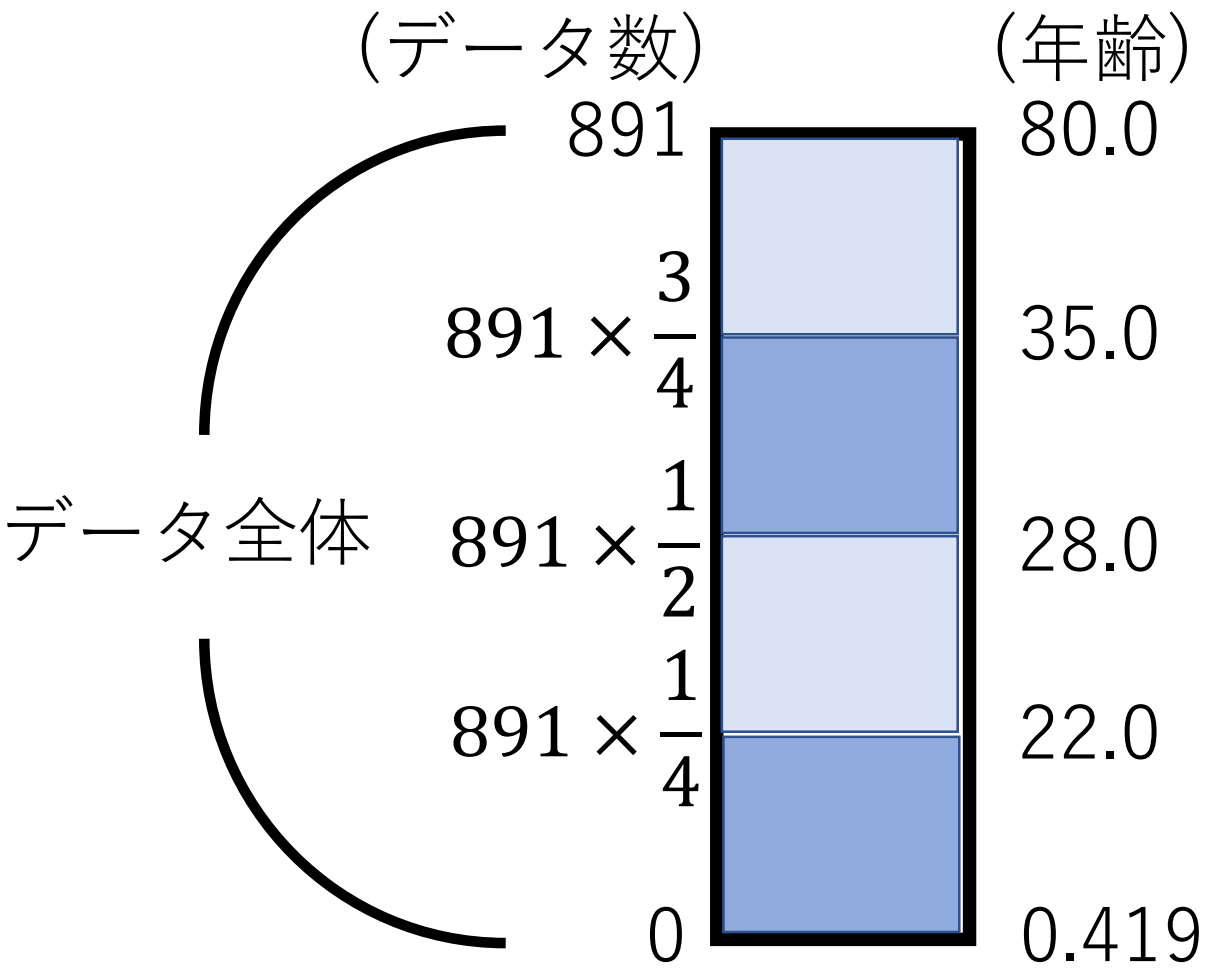
- 新しい特徴量を追加する手法

- データセットに存在する既存の特徴量に，新しく特徴量を増やすことで，機械学習モデルの予測精度を向上させる．

アイデア１：年齢の階級

年齢で4つに分けてみる！

アイデア１：年齢の階級　四分位数



35.0　＝ 第三四分位数

28.0　＝ 第二四分位数 ＝ 中央値 ≠ 平均値

22.0　＝ 第一四分位数

アイデア１：年齢の階級　qcut()

- Pandasのqcut()を用いて四分位数で分割

```
train_set['Age_Cat'] = pd.qcut(train_set['Age'], 4)
```

アイデア１：年齢の階級　分割結果

- 分割した4階級別に生存率を求める

```
AgePlot = train_set['Survived'].groupby(train_set['Age_Cat']).mean()
AgePlot
```

```
Age_Cat
(0.419, 22.0]    0.424242
(22.0, 28.0]     0.331169
(28.0, 35.0]     0.437037
(35.0, 80.0]     0.382488
Name: Survived, dtype: float64
```

階級ごとの生存率に特徴がみられる

➡新しい特徴量として採用