

実線！機械学習のプロセス徹底 底解説 -Part3 ステップ1-

16T4063F 結城洸太

ステップ1 データセットから課題を定義

- データを収集する

(今回は、タイタニック号データが用意されているため、このステップは省く)



- データセットから課題を定義

→ビジネスゴールを定義して、最適化すべきターゲットを明確化にする

「課題」と「評価指標」を考える

タイタニック号データ

データ個数	訓練データ：891
	テストデータ：418
発行元	Kaggle公式
特徴量	PassengerId：識別ユニークID
	Pclass：旅客等級
	Name：名前
	Sex：性別
	Age：年齢
	SibSp：同乗者の兄弟・配偶者の数
	Parch：同乗者の親・子供の数
	Ticket：チケット番号
	Fare：料金
	Cabin：客室番号
	Embarked：出発港
	Survived：0 (落命)
Survived：1 (生存)	

今回の課題と評価指標

- 課題 …乗船者のデータを基に、事故後にどれくらい生還できたかを、機械学習アルゴリズムで予測する
- 評価指標 …「課題」に対する評価方法。今回は初歩的な「正解率」を採用

最初の段階で、ゴールをよく考えてからプロジェクトを始めることが重要