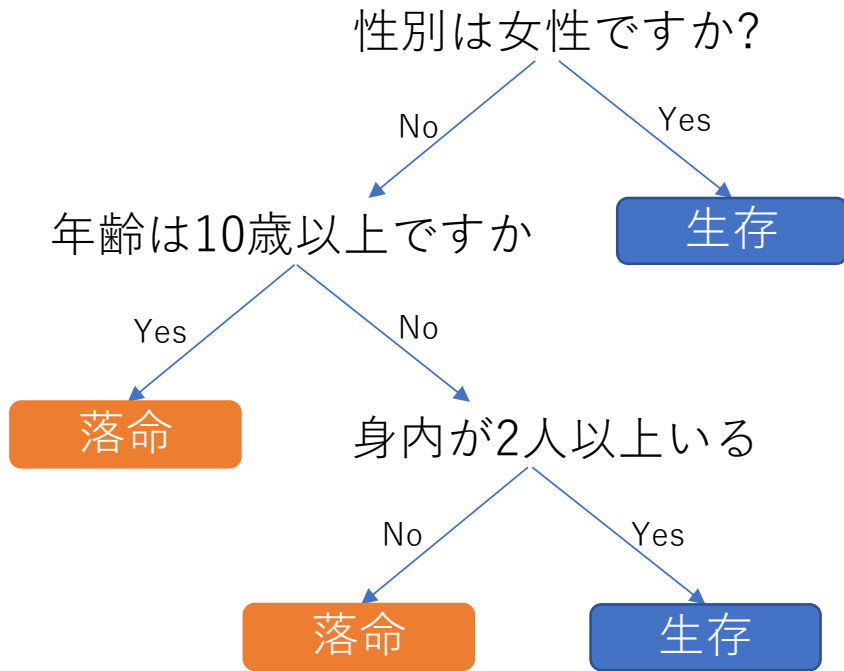


「機械学習プロセス徹底理解」 Step6 ランダムフォレスト

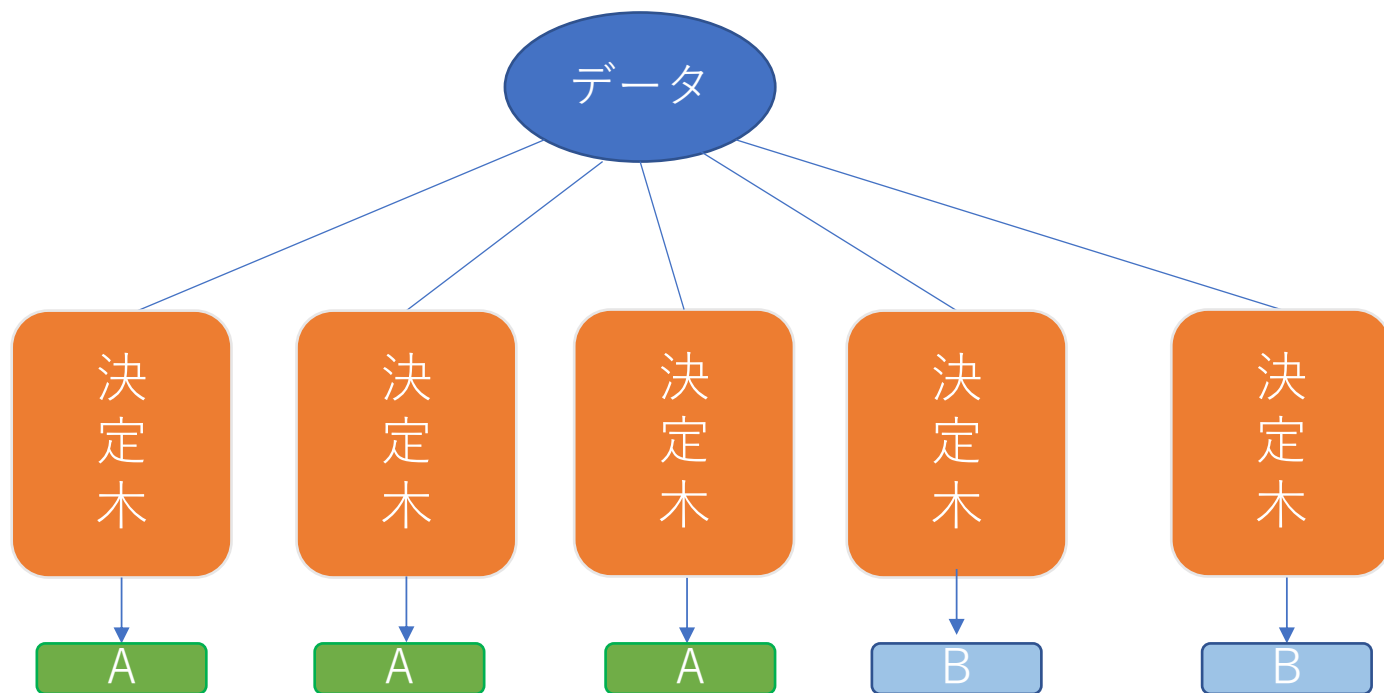
河野慎司

決定木



データに対して条件を定義していき、その一つひとつの条件に従って分類していく方法。

ランダムフォレスト



この場合, A と予想する

複数の決定木モデルを構築し, それらを1つにまとめて, より精度の高い予測を行う手法のこと.

それぞれ木ごとに異なる特徴量を選択している.

ランダムフォレスト

```
from sklearn.ensemble import RandomForestClassifier
```

```
RF = RandomForestClassifier()
```

```
RF.fit(X_train, y_train)
```

```
y_pred = RF.predict(test_set)
```

Kaggleへの投稿

この予測結果がどの程度の精度なのか、Kaggle上で答え合わせ。
予測結果をcsvで書き出す。

```
my_solution = pd.DataFrame({'PassengerId': PassengerID,  
                             'Survived': y_pred})  
my_solution = my_solution.set_index('PassengerId')  
my_solution.to_csv('submit.csv')
```

Kaggleへの投稿

<https://www.kaggle.com/c/titanic/submit> のページで書き出した submit.csv をアップロードする.

Your most recent submission				
Name	Submitted	Wait time	Execution time	Score
submit.csv	just now	0 seconds	0 seconds	0.77511
Complete				
Jump to your position on the leaderboard ▾				

約77%の生存を正しく予想できた