



dhavalsays updated decision tree tutorial

History

1 contributor

562 lines (562 sloc) | 14 KB

In [112... `import pandas as pd`

In [113... `df = pd.read_csv("titanic.csv")`
`df.head()`

Out[113...

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarke
--	-------------	----------	--------	------	-----	-----	-------	-------	--------	------	-------	---------

0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	

In [114... `df.drop(['PassengerId', 'Name', 'SibSp', 'Parch', 'Ticket', 'Cabin', 'Embarked'], axis='columns',`

In [115... `df.head()`

Out[115...

	Survived	Pclass	Sex	Age	Fare
--	----------	--------	-----	-----	------

0	0	3	male	22.0	7.2500
1	1	1	female	38.0	71.2833
2	1	3	female	26.0	7.9250
3	1	1	female	35.0	53.1000
4	0	3	male	35.0	8.0500

In [116... `inputs = df.drop('Survived', axis='columns')`
`target = df.Survived`

In [117... `inputs.Sex = inputs.Sex.map({'male': 1, 'female': 2})`

In [118... `inputs.Age[:10]`

Out[118... 0 22.0
1 38.0

```
2    26.0
3    35.0
4    35.0
5     NaN
6    54.0
7     2.0
8    27.0
9    14.0
```

Name: Age, dtype: float64

```
In [119... inputs.Age = inputs.Age.fillna(inputs.Age.mean())
```

```
In [120... inputs.head()
```

```
Out[120...
   Pclass  Sex  Age  Fare
0        3    1  22.0   7.2500
1        1    2  38.0  71.2833
2        3    2  26.0   7.9250
3        1    2  35.0  53.1000
4        3    1  35.0   8.0500
```

```
In [121... from sklearn.model_selection import train_test_split
```

```
In [122... X_train, X_test, y_train, y_test = train_test_split(inputs,target,test_size=0.2)
```

```
In [123... len(X_train)
```

```
Out[123... 712
```

```
In [124... len(X_test)
```

```
Out[124... 179
```

```
In [125... from sklearn import tree
model = tree.DecisionTreeClassifier()
```