

Aprendizado de Máquina e Deep Learning

Implementação da Árvore de Decisão

Prof. Dr. Thiago Meirelles Ventura

Implementação

- Uso da biblioteca scikit-learn

```
from sklearn.tree import DecisionTreeClassifier
```

Implementação

- Carregamento dos dados

```
df = pd.read_csv('emprego.csv')
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
0	1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.0	Mkt&HR	58.80	Placed	270000.0
1	2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed	200000.0
2	3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75.0	Mkt&Fin	57.80	Placed	250000.0
3	4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66.0	Mkt&HR	59.43	Not Placed	NaN
4	5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	Mkt&Fin	55.50	Placed	425000.0

Implementação

- Transformação dos dados

```
df['area'] = pd.factorize( df['hsc_s'] )[0]  
df['sexo'] = pd.factorize( df['gender'] )[0]  
df['experiencia'] = pd.factorize( df['workex'] )[0]
```

Implementação

- Seleção dos atributos e preparação dos conjuntos de dados

```
x = df[['area', 'sexo', 'experiencia']]  
y = df['status']
```

```
X_train, X_test, [...] = train_test_split([...])
```

Implementação

- Construção do modelo e treinamento

```
from sklearn.tree import DecisionTreeClassifier
```

```
arvore = DecisionTreeClassifier()
```

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

Implementação

- Avaliação do modelo

```
predicao = arvore.predict(X_test)
```

```
print(confusion_matrix(y_test,predicao))
```

```
print(classification_report(y_test,predicao))
```

Implementação

- Desenho da árvore

```
from sklearn import tree
```

```
features = list(x.columns)
```

```
fig, axes = plt.subplots(nrows=1, ncols=1, figsize=(20,20))
```

```
tree.plot_tree(arvore, feature_names=features, class_names=arvore.classes_)
```


Exercício 6

- Crie um modelo de Árvore de Decisão
- Aproveite da classe `DecisionTreeClassifier`
- Use o conjunto de dados “emprego.csv”

Implementação

script arvoredecisao.ipynb

Implementação

- Teste com transformação de coluna por categoria

```
df = pd.get_dummies(df, columns=['hsc_s'])
```

```
x = df[['hsc_s_Arts', 'hsc_s_Commerce',  
        'hsc_s_Science', 'sexo', 'experiençia']]
```

Implementação

script arvoredecisao.ipynb