https://chrisalbon.com/code/machine_learning/feature_selection/drop_highly_correlated_features/

In [1]:
```
# Load libraries
import pandas as pd
import numpy as np
```

In [2]:
```
# Create feature matrix with two highly correlated features
X = np.array([[1, 1, 1],
              [2, 2, 0],
              [3, 3, 1],
              [4, 4, 0],
              [5, 5, 1],
              [6, 6, 0],
              [7, 7, 1],
              [8, 7, 0],
              [9, 7, 1]])

# Convert feature matrix into DataFrame
df = pd.DataFrame(X)

# View the data frame
df
```

Out[2]:

	0	1	2
0	1	1	1
1	2	2	0
2	3	3	1
3	4	4	0
4	5	5	1
5	6	6	0
6	7	7	1
7	8	7	0
8	9	7	1

In [3]:
```
# Create correlation matrix
corr_matrix = df.corr().abs()
```

In [4]:
```
# Select upper triangle of correlation matrix
upper = corr_matrix.where(np.triu(np.ones(corr_matrix.shape), k=1).astype(bool))
```

In [5]:
```
# Find index of feature columns with correlation greater than 0.95
to_drop = [column for column in upper.columns if any(upper[column] > 0.95)]
```

In [6]:
```
to_drop
```

Out[6]:
```
[1]
```

In [7]:
```
# Drop features
df.drop(df[to_drop], axis=1)
```

Out[7]:

	0	2
0	1	1
1	2	0
2	3	1
3	4	0
4	5	1
5	6	0
6	7	1
7	8	0
8	9	1