Dimensionality Reduction

Dimensionality Reduction

- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - PCA vs LDA
- Feature Selection
  - Filter Method
    - Variance Threshold
    - Chi-Squared Test
    - Correlation Threshold
  - Wrapper Method
    - Backward Elimination
    - Recursive Feature Elimination (RFE)
  - Embedded Method
- Final thoughts

Dimensionality Reduction

- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - PCA vs LDA
- Feature Selection
  - Filter Method
    - Variance Threshold
    - Chi-Squared Test
    - Correlation Threshold
  - Wrapper Method
    - Backward Elimination
    - Recursive Feature Elimination (RFE)
  - Embedded Method
- Final thoughts

Why to apply dimensionality reduction?

- Curse of dimensionality:
  - Increases the risk of overfitting
  - Reduces the speed
  - Reduces the accuracy
  - Reduces explainability
- Visualisation
- Types of dimensionality reduction:
  - Feature extraction
  - Feature selection

Dimensionality Reduction

- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
    - Principal Component Analysis (PCA)
    - Linear Discriminant Analysis (LDA)
    - PCA vs LDA
- Feature Selection
    - Filter Method
        - Variance Threshold
        - Chi-Squared Test
        - Correlation Threshold
    - Wrapper Method
        - Backward Elimination
        - Recursive Feature Elimination (RFE)
    - Embedded Method
- Final thoughts

Models overview

- Feature Selection
- Feature Extraction

Dimensionality Reduction

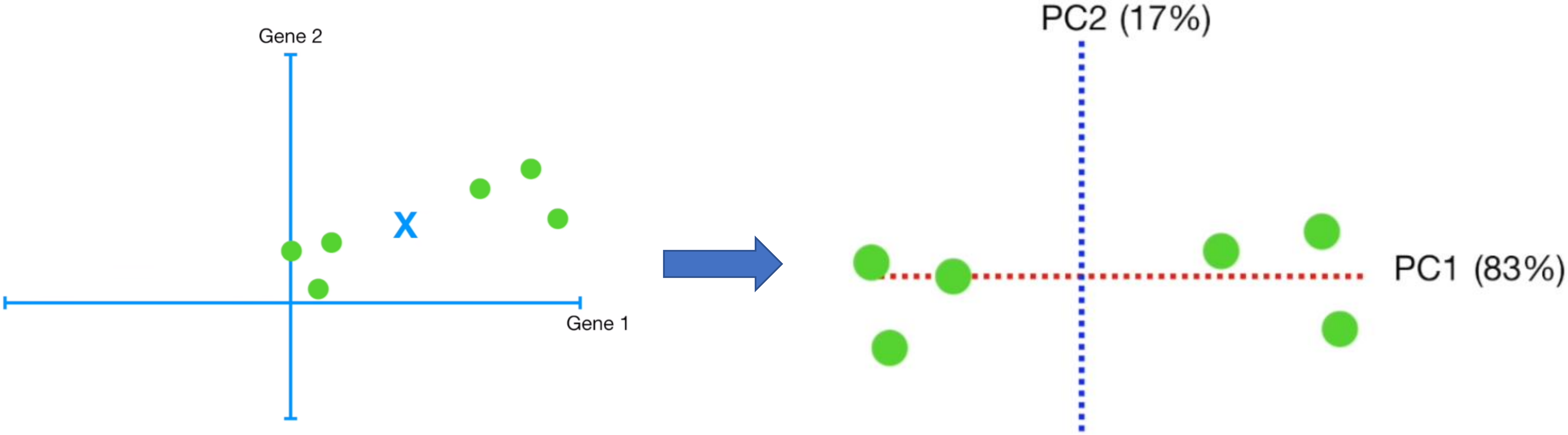
- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
    - Principal Component Analysis (PCA)
    - Linear Discriminant Analysis (LDA)
    - PCA vs LDA
- Feature Selection
    - Filter Method
        - Variance Threshold
        - Chi-Squared Test
        - Correlation Threshold
    - Wrapper Method
        - Backward Elimination
        - Recursive Feature Elimination (RFE)
    - Embedded Method
- Final thoughts

Principal Component Analysis (PCA)

- Unsupervised
- Creates linear combinations
- New features are orthogonal
- Maximizes variance by considering the variance of each feature
- Features are ranked in order of their variance
- Needs feature scaling
- Only works for continuous features

- Kernel PCA for non-linear separable datasets

Principal Component Analysis (PCA)



https://www.youtube.com/watch?v=FgakZw6K1QQ

Principal Component Analysis (PCA)

	Cumulative Variance Ratio	Explained Variance Ratio
0	0.449303	0.449303
1	0.639841	0.190538
2	0.728390	0.088549
3	0.793421	0.065031
4	0.847024	0.053603
5	0.887960	0.040937
6	0.910797	0.022837
7	0.928059	0.017262
8	0.941753	0.013694
9	0.953757	0.012004
10	0.964381	0.010623
11	0.972419	0.008038
12	0.979418	0.006999
13	0.984170	0.004751
14	0.987257	0.003087

https://www.youtube.com/watch?v=FgakZw6K1QQ

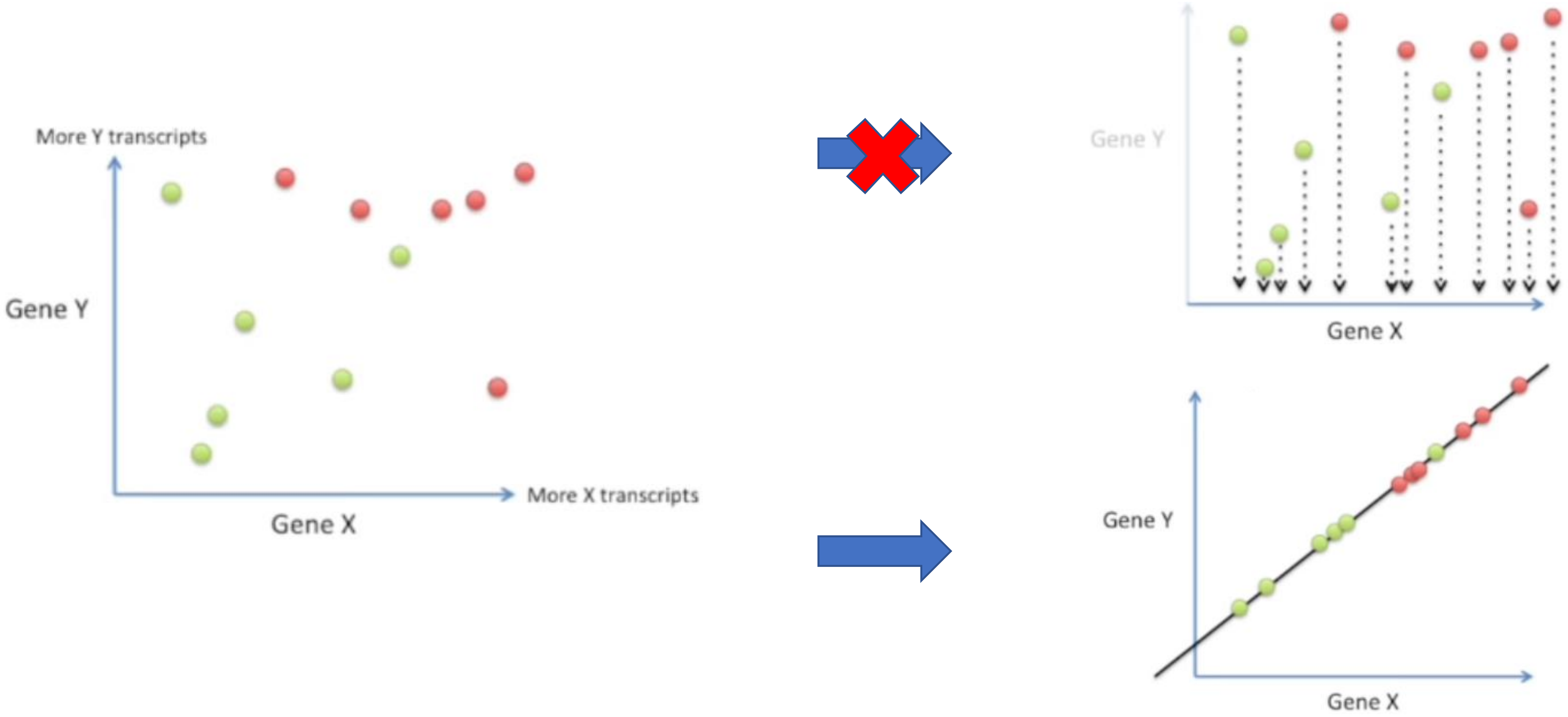
Dimensionality Reduction

- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
    - Principal Component Analysis (PCA)
    - Linear Discriminant Analysis (LDA)
    - PCA vs LDA
- Feature Selection
    - Filter Method
        - Variance Threshold
        - Chi-Squared Test
        - Correlation Threshold
    - Wrapper Method
        - Backward Elimination
        - Recursive Feature Elimination (RFE)
    - Embedded Method
- Final thoughts

Linear Discriminant Analysis (LDA)

- Supervised
- Creates linear combinations
- New features are orthogonal
- Maximizes separability between classes by considering the information of classes
- Features are ranked in order of their separability between classes
- Doesn’t need feature scaling
- Only works for continuous features

Linear Discriminant Analysis (LDA)



https://www.youtube.com/watch?v=azXCzI57Yfc

Linear Discriminant Analysis (LDA)

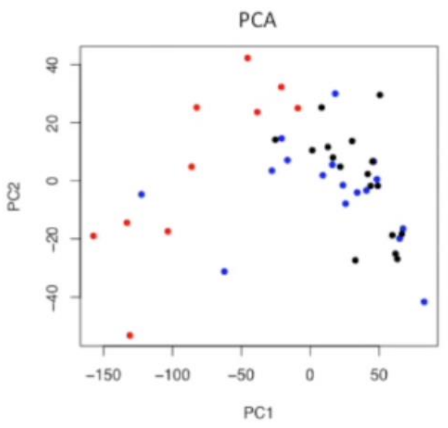
	Cumulative Variance Ratio	Explained Variance Ratio
0	0.667519	0.667519
1	1.000000	0.332481

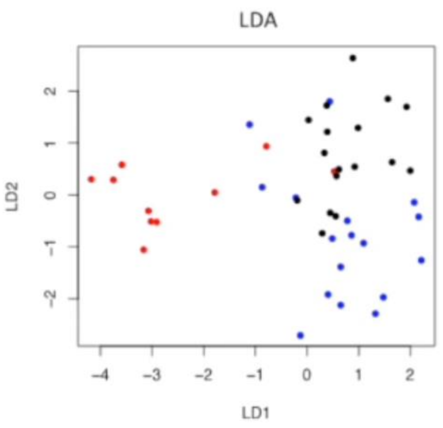
https://www.youtube.com/watch?v=azXCzI57Yfc

Dimensionality Reduction

- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - PCA vs LDA
- Feature Selection
  - Filter Method
    - Variance Threshold
    - Chi-Squared Test
    - Correlation Threshold
  - Wrapper Method
    - Backward Elimination
    - Recursive Feature Elimination (RFE)
  - Embedded Method
- Final thoughts

PCA vs LDA





- Unsupervised
- Creates linear combinations
- New features are orthogonal
- Maximizes variance by considering the variance of each feature
- Features are ranked in order of their variance
- Only works for continuous features
- Needs feature scaling

- Kernel PCA for non-linear separable datasets

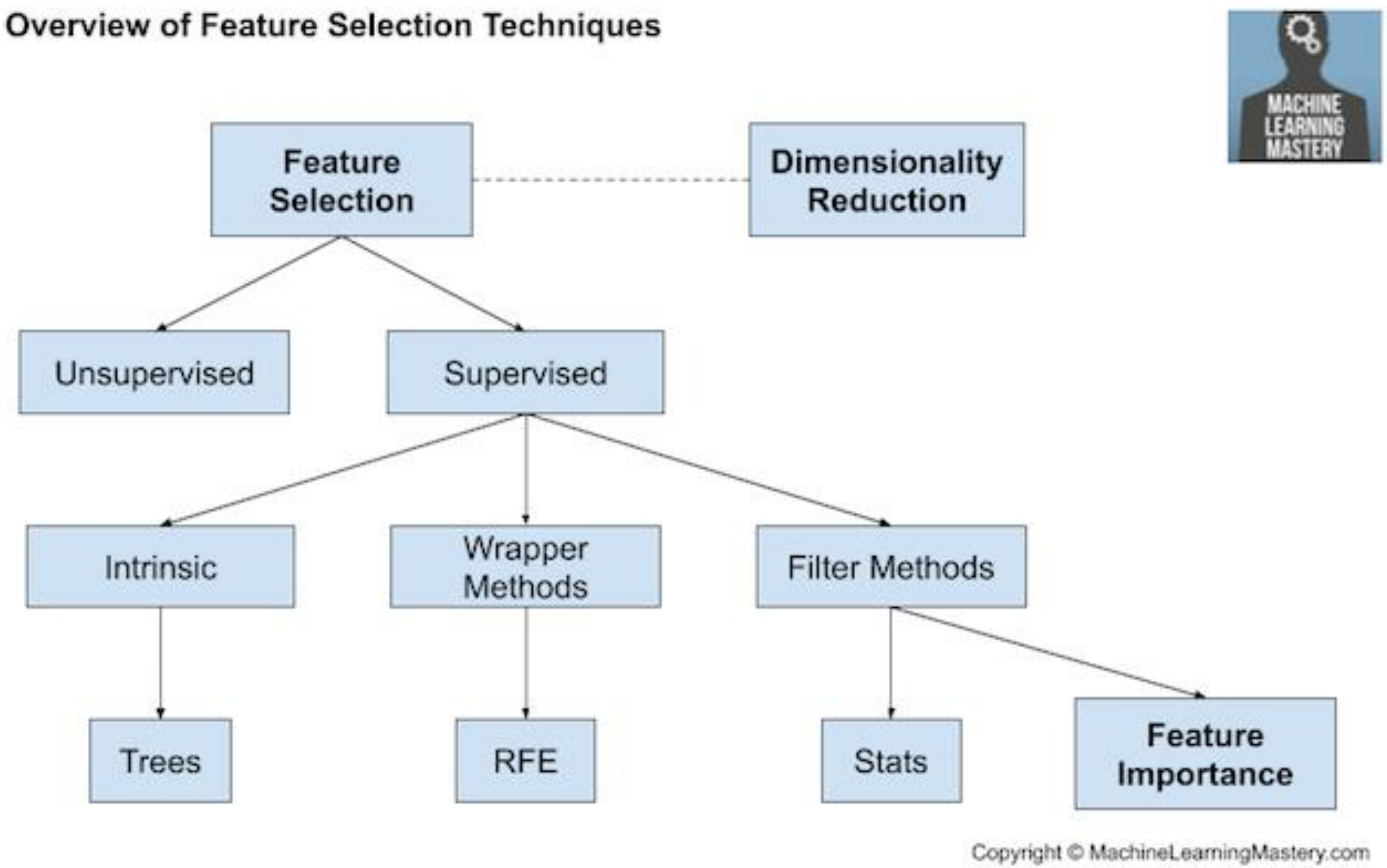
- Supervised
- Creates linear combinations
- New features are orthogonal
- Maximizes separability between classes by considering the information of classes
- Features are ranked in order of their separability between classes
- Only works for continuous features
- Doesn’t need feature scaling

https://www.youtube.com/watch?v=azXCzI57Yfc

Dimensionality Reduction

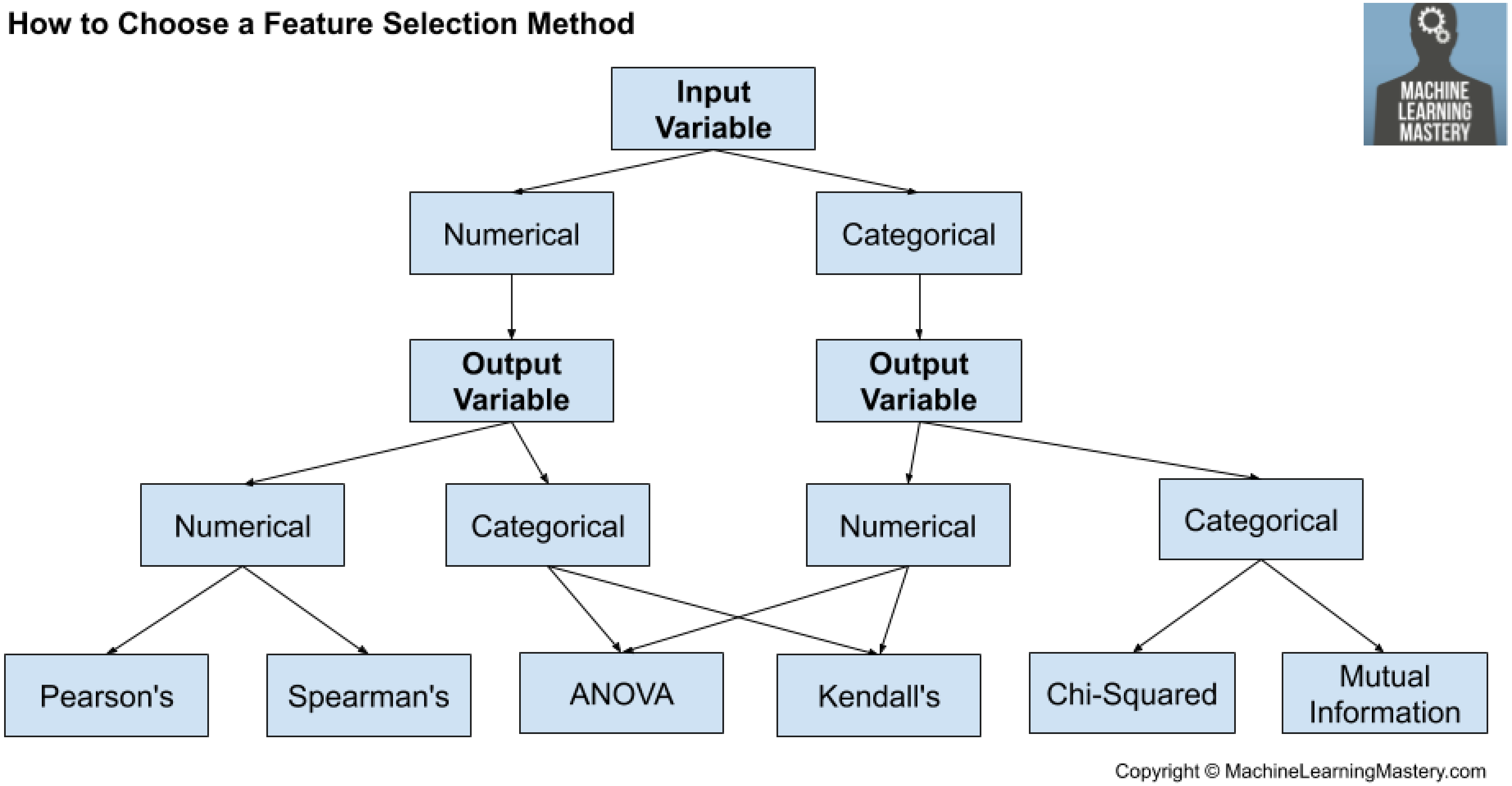
- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - PCA vs LDA
- Feature Selection
  - Filter Method
    - Variance Threshold
    - Chi-Squared Test
    - Correlation Threshold
  - Wrapper Method
    - Backward Elimination
    - Recursive Feature Elimination (RFE)
  - Embedded Method
- Final thoughts

Feature Selection



https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/

Feature Selection



https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/

Dimensionality Reduction

- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - PCA vs LDA
- Feature Selection
  - Filter Method
    - Variance Threshold
    - Chi-Squared Test
    - Correlation Threshold
  - Wrapper Method
    - Backward Elimination
    - Recursive Feature Elimination (RFE)
  - Embedded Method
- Final thoughts

Variance Threshold

- Unsupervised
- Removes features whose values don’t change much from observation to observation
- Needs feature scaling (but it can’t be Standard Scaler)
- Works for both continuous and one hot encoded features

Variance Threshold

Toy example:

```
from sklearn.feature_selection import VarianceThreshold
df_selected, df_removed = variance_threshold_selector(X_train, 0.01)
```

	Var 1	Var 2	Var 3
0	1.00	1	5
1	1.01	2	5
2	1.02	3	5
3	1.03	4	5

	Var 1	Var 2	Var 3
1	1.01	2	5
0	1.00	1	5
2	1.02	3	5

	Var 1	Var 2	Var 3
3	1.03	4	5

	Var 1	Var 2	Var 3
0	0.577331	0.534522	0.57735
1	0.571615	0.267261	0.57735
2	0.583048	0.801784	0.57735

```
df_selected
```

	Var 2
0	0.534522
1	0.267261
2	0.801784

```
df_removed
```

	Var 1	Var 3
0	0.577331	0.57735
1	0.571615	0.57735
2	0.583048	0.57735

Dimensionality Reduction

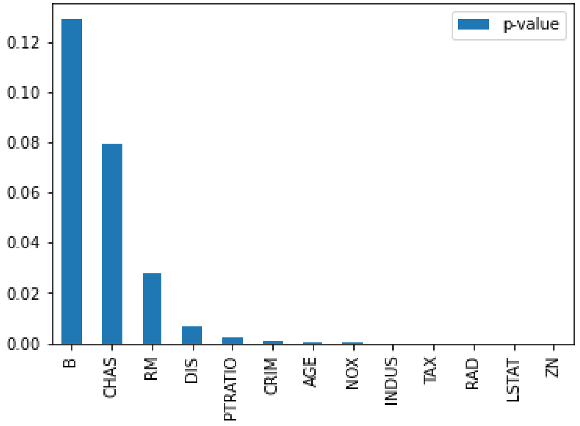
- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
    - Principal Component Analysis (PCA)
    - Linear Discriminant Analysis (LDA)
    - PCA vs LDA
- Feature Selection
    - Filter Method
        - Variance Threshold
        - Chi-Squared Test
        - Correlation Threshold
    - Wrapper Method
        - Backward Elimination
        - Recursive Feature Elimination (RFE)
    - Embedded Method
- Final thoughts

Chi-Squared Test

- Supervised
- Removes features whose that don’t effect the dependent variable
- Needs feature scaling (but it can’t be Standard Scaler)
- Only works for one hot encoded features
- Only works for classification

Chi-Squared Test

	p-value
B	1.290574e-01
CHAS	7.971755e-02
RM	2.767903e-02
DIS	6.525865e-03
PTRATIO	2.472616e-03
CRIM	1.270929e-03
AGE	1.594477e-04
NOX	3.522787e-05
INDUS	7.555366e-06
TAX	4.143725e-07
RAD	4.005216e-07
LSTAT	8.926390e-08
ZN	2.879796e-08



Dimensionality Reduction

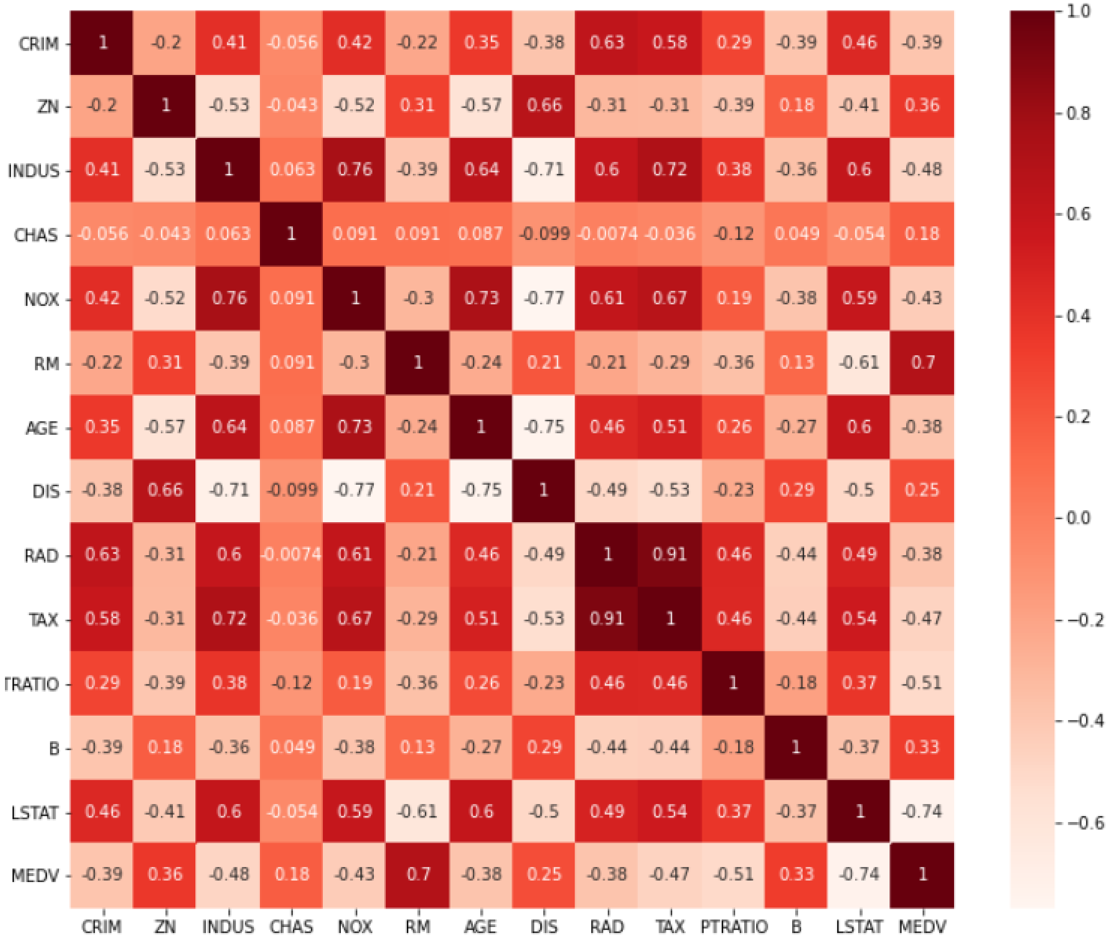
- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - PCA vs LDA
- Feature Selection
  - Filter Method
    - Variance Threshold
    - Chi-Squared Test
    - Correlation Threshold
  - Wrapper Method
    - Backward Elimination
    - Recursive Feature Elimination (RFE)
  - Embedded Method
- Final thoughts

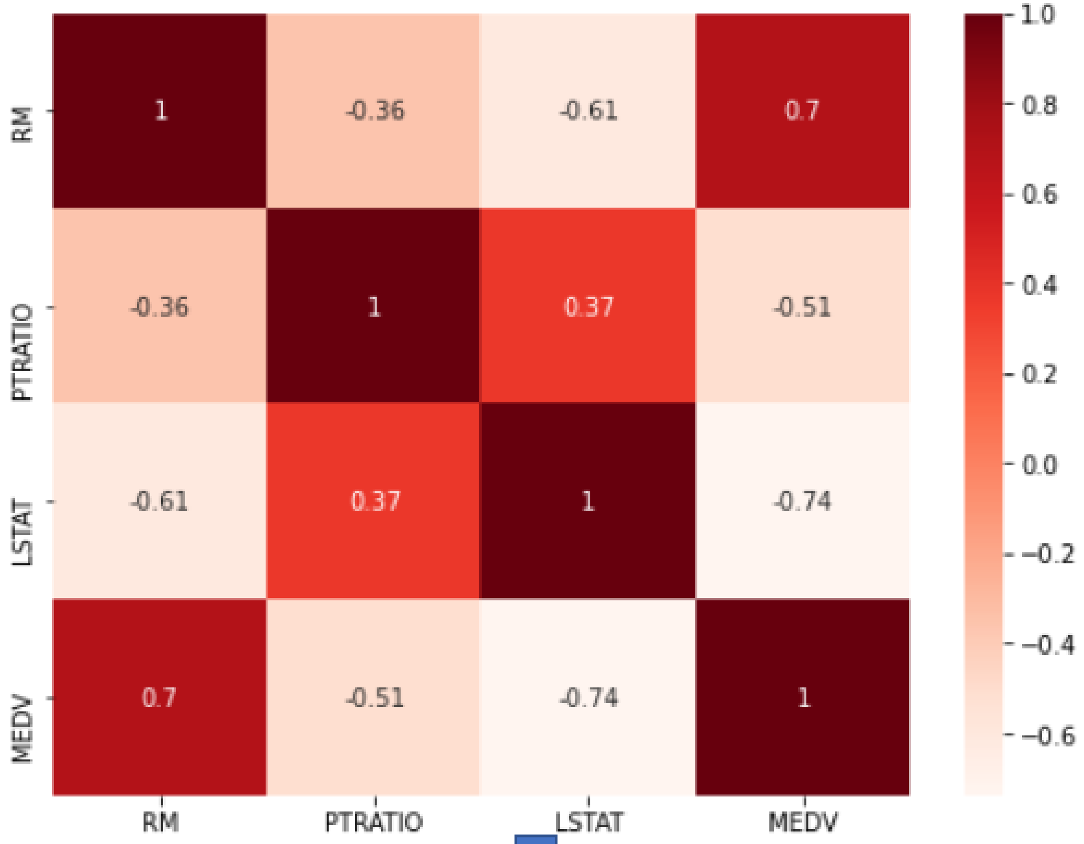
Correlation threshold

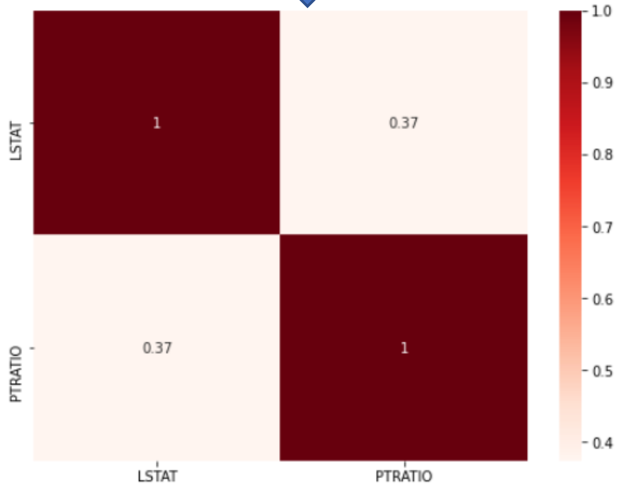
- Supervised/Unsupervised
- Removes features that are not highly correlated with the dependent variable
- Removes features that are highly correlated with each other
- Needs feature scaling
- “Works” for both continuous and one hot encoded features
- Works for both regression and classification
- Specific for categorical features:

  https://towardsdatascience.com/the-search-for-categorical-correlation-a1cf7f1888c9

Correlation threshold








```
relevant_features = cor_target[cor_target > 0.5]
```

Dimensionality Reduction

- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - PCA vs LDA
- Feature Selection
  - Filter Method
    - Variance Threshold
    - Chi-Squared Test
    - Correlation Threshold
  - Wrapper Method
    - Backward Elimination
    - Recursive Feature Elimination (RFE)
  - Embedded Method
- Final thoughts

Backward Elimination

- Supervised
- Start with all features and remove one at a time using the p-value
- Not recommended
- Needs feature scaling
- Works for both continuous and one hot encoded features
- Works for both regression and classification
- Model example: Ordinary Least Squares

Backward Elimination

```
['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX', 'PTRATIO', 'B', 'LSTAT']
```

```
CRIM       1.086810e-03
ZN         7.781097e-04
INDUS      7.382881e-01
CHAS       1.925030e-03
NOX        4.245644e-06
RM         1.979441e-18
AGE        9.582293e-01
DIS        6.013491e-13
RAD        5.070529e-06
TAX        1.111637e-03
PTRATIO    1.308835e-12
B          5.728592e-04
LSTAT      7.776912e-23
dtype: float64
0.9582293092057567
AGE
```

```
CRIM       1.074747e-03
ZN         7.193806e-04
INDUS      7.379887e-01
CHAS       1.862634e-03
NOX        1.967110e-06
RM         3.365945e-19
DIS        5.027955e-14
RAD        4.750539e-06
TAX        1.099120e-03
PTRATIO    1.099178e-12
B          5.444689e-04
LSTAT      2.569688e-25
dtype: float64
0.7379887092915616
INDUS
```

```
CRIM       1.010438e-03
ZN         7.542759e-04
CHAS       1.551469e-03
NOX        1.209413e-06
RM         2.889779e-19
DIS        6.837043e-15
RAD        2.996799e-06
TAX        5.214237e-04
PTRATIO    9.235063e-13
B          5.565743e-04
LSTAT      2.140586e-25
dtype: float64
0.0015514692639118284
CHAS
```

```
['CRIM', 'ZN', 'CHAS', 'NOX', 'RM', 'DIS', 'RAD', 'TAX', 'PTRATIO', 'B', 'LSTAT']
```

Dimensionality Reduction

- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - PCA vs LDA
- Feature Selection
  - Filter Method
    - Variance Threshold
    - Chi-Squared Test
    - Correlation Threshold
  - Wrapper Method
    - Backward Elimination
    - Recursive Feature Elimination (RFE)
  - Embedded Method
- Final thoughts

Recursive Feature Elimination (RFE)

- Supervised
- Recursively prunes the number of features until the desired number of features is reached
- Needs feature scaling
- Works for both continuous and one hot encoded features
- Works for both regression and classification
- Model examples: any model that has either coef_ or featyre_importances_ attribute

Recursive Feature Elimination (RFE)

```
Possible optimum number of features: 11
Score with 11 features: 0.675177
```

	Score
1	0.521805
3	0.598559
2	0.606979
4	0.618632
6	0.628443
5	0.630139
7	0.641448
8	0.647980
9	0.655982
10	0.663581
12	0.673230
13	0.673383
11	0.675177

Dimensionality Reduction

- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - PCA vs LDA
- Feature Selection
  - Filter Method
    - Variance Threshold
    - Chi-Squared Test
    - Correlation Threshold
  - Wrapper Method
    - Backward Elimination
    - Recursive Feature Elimination (RFE)
  - Embedded Method
- Final thoughts

Embedded Method

- Supervised machine learning algorithms that perform feature selection automatically
- Needs feature scaling
- Works for both continuous and one hot encoded features
- Works for both regression and classification

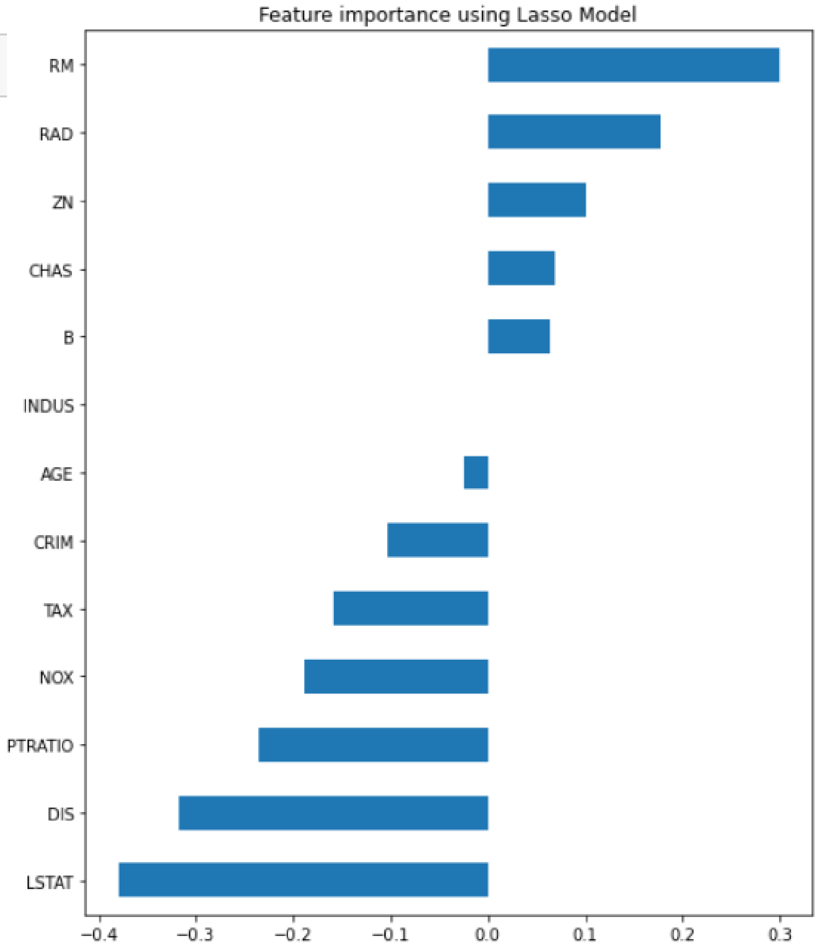
Embedded Method

```
coef = pd.Series(reg.coef_, index = df.drop('MEDV', axis=1).columns)
print("Lasso picked", str(sum(coef != 0)), "variables and eliminated the other", str(sum(coef == 0)), "variables")
```

```
Lasso picked 12 variables and eliminated the other 1 variables
```

```
coef.sort_values()
```

```
LSTAT     -0.379874
DIS       -0.317952
PTRATIO   -0.236031
NOX       -0.188779
TAX       -0.159086
CRIM      -0.102860
AGE       -0.024037
INDUS     -0.000000
B          0.063249
CHAS       0.068925
ZN         0.100789
RAD        0.177640
RM         0.298996
dtype: float64
```



Dimensionality Reduction

- Why to apply dimensionality reduction?
- Models overview
- Feature Extraction
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - PCA vs LDA
- Feature Selection
  - Filter Method
    - Variance Threshold
    - Chi-Squared Test
    - Correlation Threshold
  - Wrapper Method
    - Backward Elimination
    - Recursive Feature Elimination (RFE)
  - Embedded Method
- Final thoughts

Final thoughts

- Filter method is less accurate
- Backward Elimination performs poorly
- Wrapper and Embedded methods give more accurate results but are compututionally expensive, these methods are suited when you have lesses features (about 20)

Final thoughts

	Feature	Variance	Chi-Squared	Correlation	Backward	RFE	Lasso	Total
1	TAX	True	True	True	True	True	True	6
2	PTRATIO	True	True	True	True	True	True	6
3	LSTAT	True	True	True	True	True	True	6
4	ZN	True	True	False	True	True	True	5
5	RM	True	True	False	True	True	True	5
6	RAD	True	True	False	True	True	True	5
7	NOX	True	True	False	True	True	True	5
8	DIS	True	True	False	True	True	True	5
9	CRIM	False	True	False	True	True	True	4
10	CHAS	True	False	False	True	True	True	4
11	B	True	False	False	True	True	True	4
12	INDUS	True	True	False	False	False	True	3
13	AGE	True	True	False	False	False	False	2