LECTURER: TAI LE QUY

MACHINE LEARNING

UNSUPERVISED LEARNING AND FEATURE ENGINEERING

UNIT 5

FEATURE SELECTION

STUDY GOALS

- Describe the **different techniques** used to **select relevant features**.
- Explain how to **rank features** according to certain relevant **evaluation criteria**.
- Select the best features in order to **maximize the performances** and **avoid overfitting** of the learned model.

EXPLAIN SIMPLY

1. Briefly describe the **three main influences** of **machine learning models' performances**.

2. Explain how **feature variance** can be used for **feature selection**.

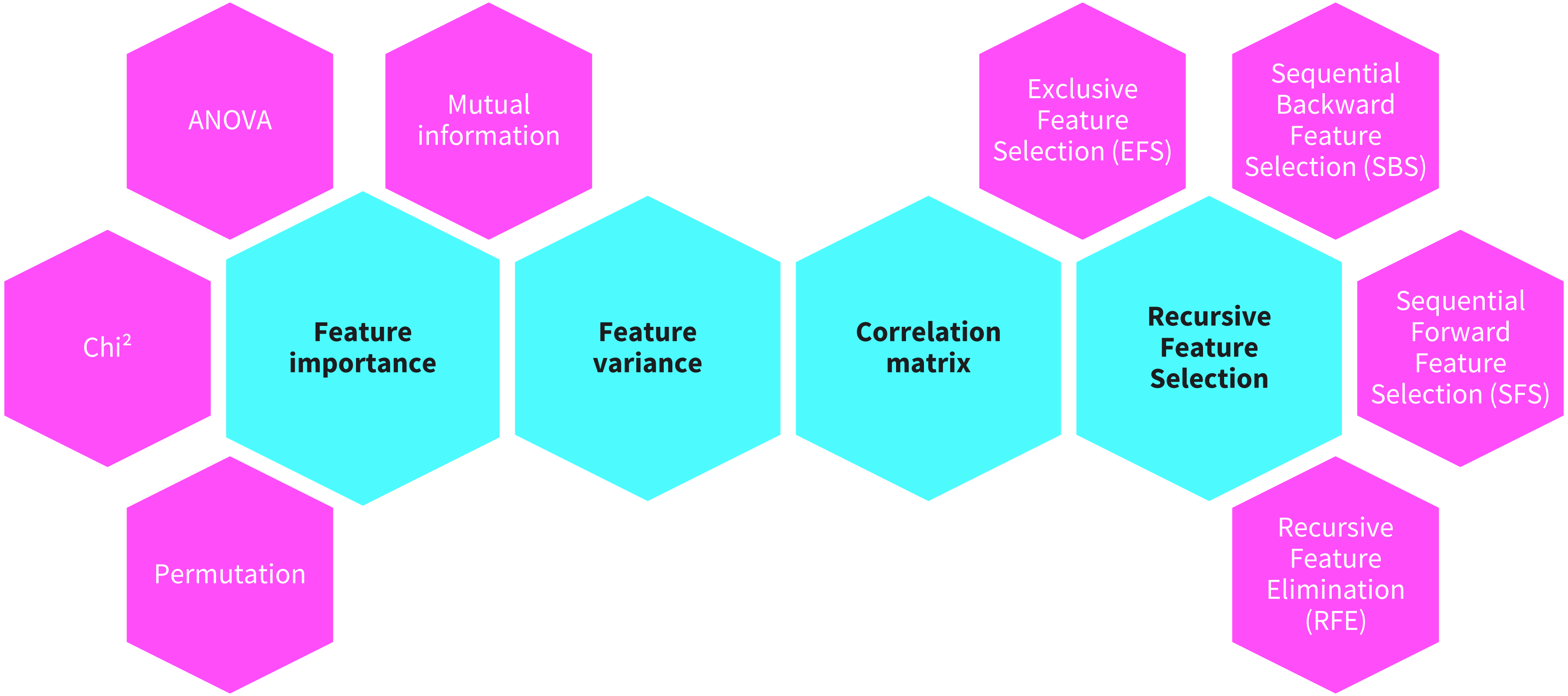
3. Describe the difference between **bottom-up** and **top-down recursive feature selection**.

- Feature selection is the process of choosing relevant features from the original features
- Feature selection can be performed manually or with algorithms
- Feature selection algorithms:
  - Filter models: features are selected by studying their characteristics using some statistical evaluation criteria
  - Wrapper models: using a certain learning approach
- Selected features are identified either by:
  - Index: the rank of a feature (e.g., a feature with rank 1 means that it is the best feature with respect to its relevance or importance)
  - Weight: the relevance of features (the higher the feature weight, the better its relevance or rank)
- Feature selection can be either univariate or multivariate
  - Univariate feature selection: each feature is independently evaluated
  - Multivariate feature selection: each feature is evaluated with respect to other features

- Feature selection algorithms:
  - Supervised Feature Selection (SFS): labels (output) of data points are available, selects discriminant features that allow separate data points to belong to different classes.
  - Unsupervised Feature Selection (UFS): labels are not available. UFS is much harder to perform than SFS because defining a feature's relevancy in the absence of output (labels) becomes challenging.
- Feature selection can be either univariate or multivariate
  - Univariate feature selection: each feature is independently evaluated
  - Multivariate feature selection: each feature is evaluated with respect to other features

UNIT CONTENT

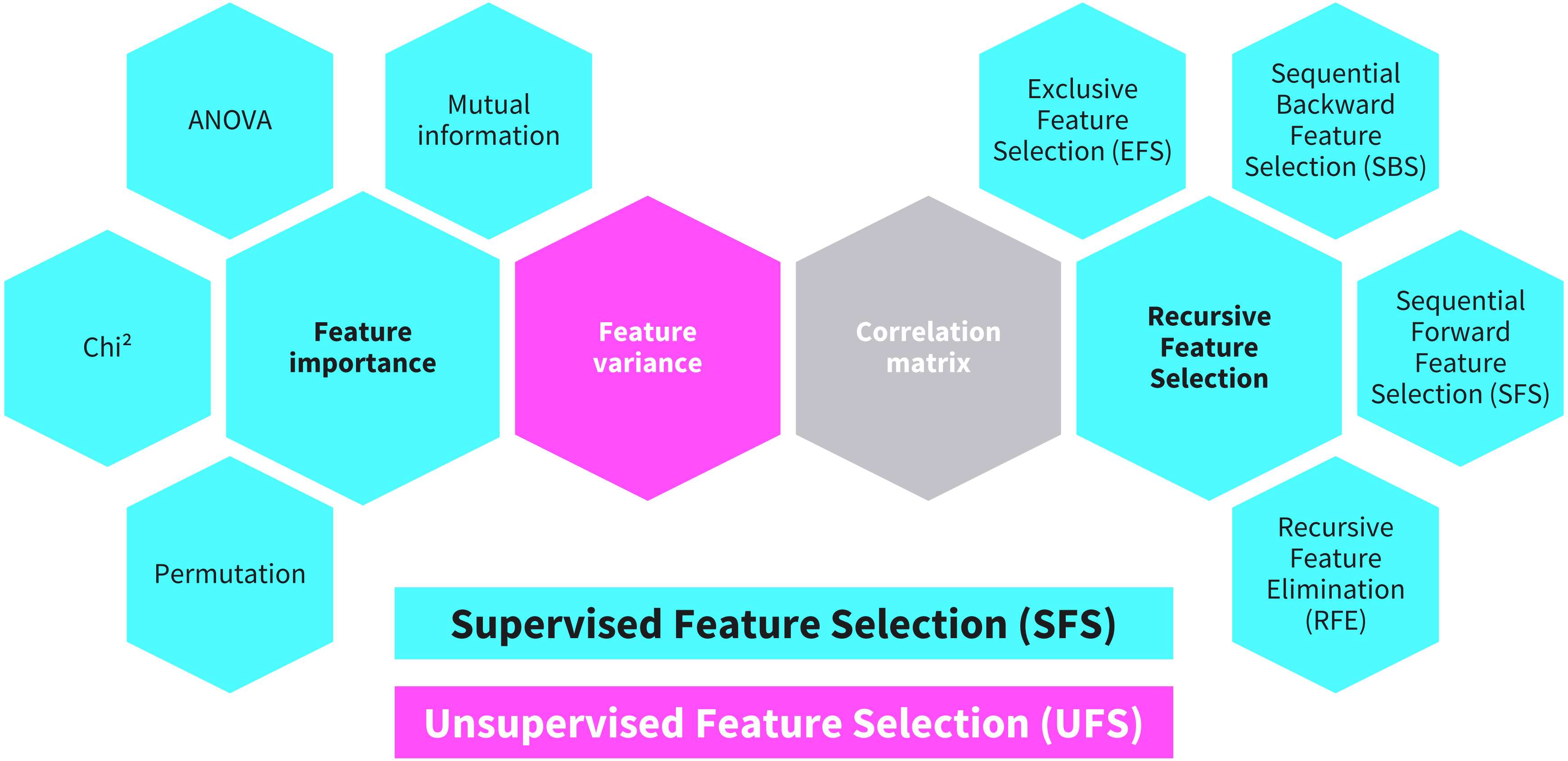
Image 1: Unit content - Feature selection



Source of the image : Müller-Kett, 2021.

UNIT CONTENT

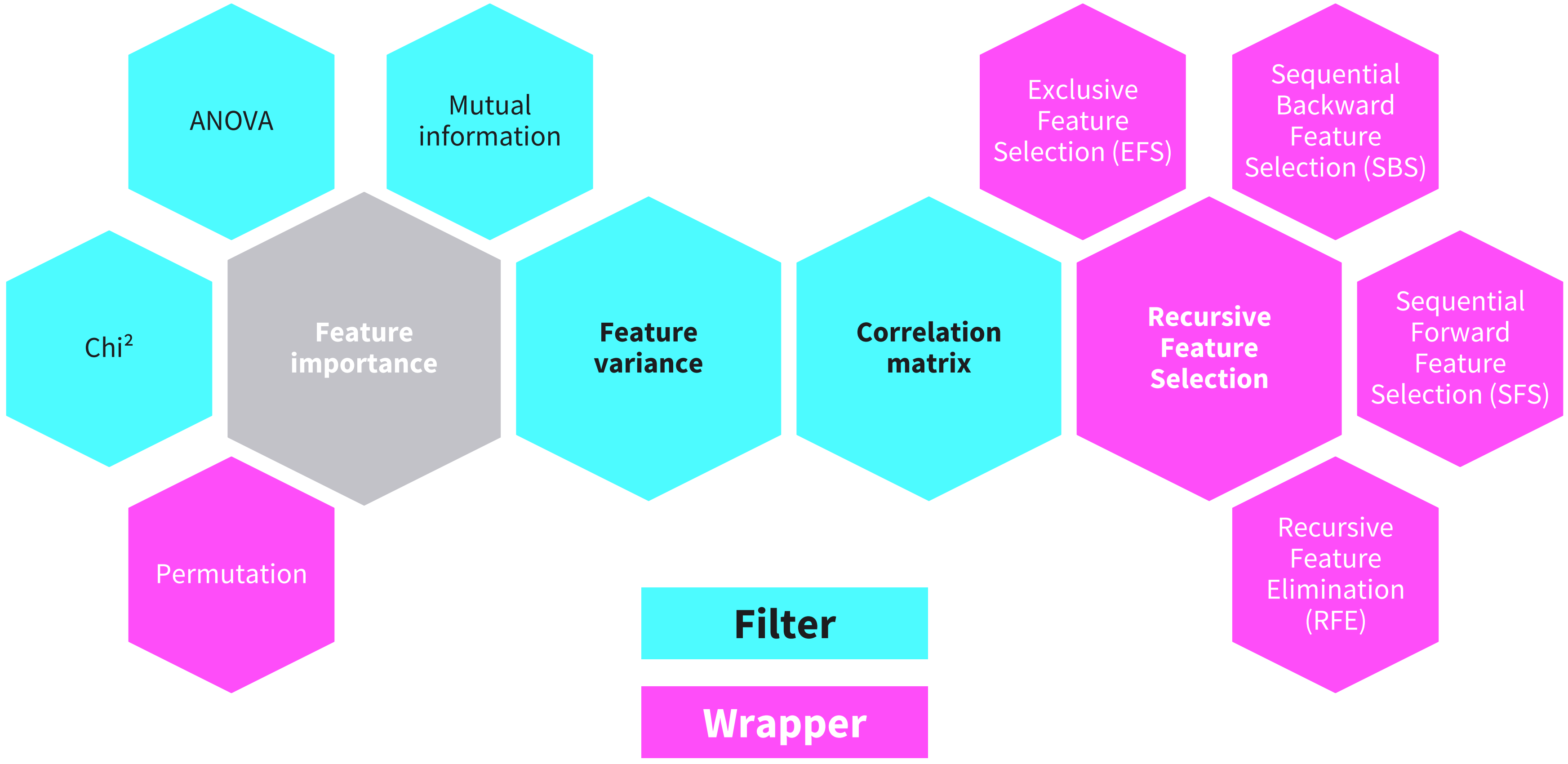
Image 1: Unit content - Feature selection



Source of the image : Müller-Kett, 2021.

UNIT CONTENT

Image 1: Unit content - Feature selection



Source of the image : Müller-Kett, 2021.

- Features can be ranked according to their relevance to the response variable (output and labels)
  - Assigning a score to features according to their contribution to the prediction of the response
  - Evaluate correlation between each feature and the response variable
    - The higher the correlation between an input feature and the response variable, the better the score or the importance of this feature in predicting the response
  - Feature importance is primarily a supervised feature selection technique.
- Methods
  - ANOVA Test
  - Chi-Square Test
  - Mutual Information

- Compares the variance of one independent feature (input feature) with one dependent feature (output feature) to assess whether they are relevant
- Formulation
  - A dataset X containing k input features (columns) and n data points (rows).
  - Let X_{jm}, j = 1, .., k, be the mean value of each feature j.
  - Let X_m be the overall mean of the dataset.
  - The sum of squares between input features (SSB)

$$SSB = \sum_{j=1}^{k} n \cdot \left(X_{jm} - X_m\right)^2$$

  - The sum of the squared differences between each data point and its corresponding mean, X_{jm}

$$\mathrm{SSE} = \sum_{j=1}^{k} \sum_{i=1}^{n} \left(X_{ji} - X_{jm}\right)^2$$

- Formulation
  - Mean square between input features (MSB) and the mean square of errors (MSE)

$$MSB = \frac{SSB}{k-1}, MSE = \frac{SSE}{n-k}$$

k – 1: the degree of freedom (df1), n - k: the degree of freedom (df2)

  - F-value used to accept or reject the null hypothesis
  - Check in F-table

$$F = \frac{MSB}{MSE}$$

- Hypothesis
  - Null hypothesis: feature has no relevance to the response (target) variable
  - Alternative hypothesis: feature has some relevance to the response variable
  - p-value probabilities: probability or chance that the data points occurred under the null hypothesis

- Hypothesis
  - The lower the p-value of a certain feature, the better its chance to have some relevance (i.e., rejecting the null hypothesis for the alternative hypothesis)
  - Choose the confidence level indicating our confidence in rejecting or accepting the null hypothesis
    - E.g., if we want to be within the confidence interval bigger than or equal to 95 percent and less than 100 percent, an **Alpha level** of 5%.

• $p > 5\%$ means “not significant,” shown as “n.s.” in graphics.
• $0.01 < p <= 5\%$ means “significant,” shown with an asterisk, *, in graphics.
• $0.001 < p <= 0.01$ means “highly significant,” shown with two asterisks, **, in graphics.
• $p <= 0.001$ means “very highly significant,” shown with three asterisks, ***, in graphics.

F-TABLE

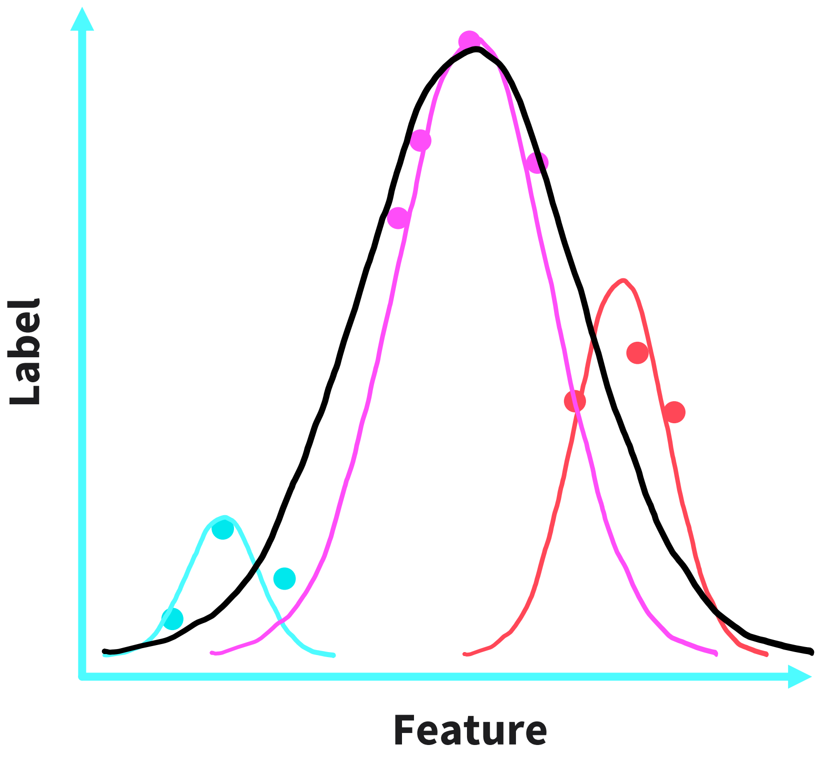
F-table of Critical Values for Significance Level = 0.05

F-table of Critical Values of α = 0.05 for F(df1, df2)																			
	DF1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
DF2=1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Source of image: https://statisticsbyjim.com/hypothesis-testing/f-

- Compares **differences** in **group means.**
- **Statistical hypothesis test**
- **Input feature**: Numeric
- **Labels**: Categorical
- Choose the features with the **lowest probability** that there is **no difference between groups** (low p-value).

Image 2: ANOVA



Source of text: Bejani et al., 2014.
Source of the image: Müller-Kett, 2023.

- Chi-square test allows the evaluation of the independence between 2 variables
- Calculated based on the difference between the observed O and expected E values for input feature with respect to each category of the response variable:

$$Chi - square = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

- Example:

Contingency table

Gen-der	Responded-Yes	Responded-No	Total
Male	O(Male/Yes)	O(Male/No)	n_{Male} = O(Male/Yes) + O(Male/No)
Female	O(Female/Yes)	O(Female/No)	n_{Female} = O(Female/Yes) + O(Female/No)
Total	n_{Yes} = O(Male/Yes) + O(Female/Yes)	n_{No} = O(Male/No) + O(Female/No)	$n = n_{Male} + n_{Female}$

- We consider the patient gender (input feature) independent from the treatment response (response variable)

$$E(Male|Yes) = n \cdot p(Male) \cdot p(Yes) = n \cdot \frac{n_{Male}}{n} \cdot \frac{n_{Yes}}{n}$$

$$E(Male|No) = n \cdot p(Male) \cdot p(No) = n \cdot \frac{n_{Male}}{n} \cdot \frac{n_{No}}{n}$$

$$E(Female|Yes) = n \cdot p(Female) \cdot p(Yes) = n \cdot \frac{n_{Female}}{n} \cdot \frac{n_{Yes}}{n}$$

$$E(Female|No) = n \cdot p(Female) \cdot p(No) = n \cdot \frac{n_{Female}}{n} \cdot \frac{n_{No}}{n}$$

- The Chi-square score

$$\begin{aligned}\text{Chi} - \text{square} \; &= \frac{(O(Male|Yes) - E(Male|Yes))^2}{E(Male|Yes)} \\ &+ \frac{(O(Male|No) - E(Male|No))^2}{E(Male|No)} \\ &+ \frac{(O(Female|Yes) - E(Female|Yes))^2}{E(Female|Yes)} \\ &+ \frac{(O(Female|No) - E(Female|No))^2}{E(Female|No)}\end{aligned}$$

CHI^2 TEST

- Chi-square test allows us to test the relationship between an input categorical feature with a categorical output or response variable
- The **null** hypothesis is that the input feature and response variable are **independent**
- Greater the difference between the observed and the expected values, more the Chi-square value → the input feature (patient gender) is dependent on the response variable (treatment response)
- Higher values of Chi-square tests indicate that the feature is more **dependent** on the response variable, i.e., it has more **importance**.

CHI-SQUARE DISTRIBUTION TABLE

Degrees of freedom (*df*)	Significance level (α) .99	.975	.95	.9	.1	.05	.025	.01
1	-------	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892

Source of image: https://www.scribbr.com/statistics/chi-square-distribution-table/

- Compares **expected and observed values** in a **contingency table**
- This table lists each **unique feature value** and **label**, as well as their **occurrences** and **all combinations**
- **Statistical hypothesis test**
- **Input feature**: Categorical
- **Labels**: Categorical
- Choose the features with the lowest probability of **not being related with the labels (low p-value)**

Table 1: Contingency table

	Cat	Dog	Total
Purrs	(cat|purrs) 678	(dog|purrs) 3	(purrs) 681
Barks	(cat|barks) 16	(dog|barks) 129	(barks) 145
Total	(cat) 694	(dog) 132	826

Source of the table: Müller-Kett, 2023.

MUTUAL INFORMATION

- Measures the amount of information gained (the reduction in the uncertainty)
- One variable or feature given a known value of another feature
  - X and Y share mutual information
- The mutual information between two variables X and Y:

$$MI(X,Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \cdot \log_2\left(\frac{p(x,y)}{p(x) \cdot p(y)}\right)$$

where p(x) and p(y) are the marginal probabilities, and p(x, y) is the joint probability

$$MI(X,Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \cdot \log_2\left(\frac{p(x,y)}{p(x) \cdot p(y)}\right)$$

- Example

X = x1	X = x1	X = x1	X = x1	X = x2	X = x2
Y = 1	Y = 2	Y = 2	Y = 2	Y = 1	Y = 2

p(x1) = 4/6, p(x2) = 2/6, p(Y=1) = 2/6, p(Y = 2) = 4/6

p(x1,Y = 1)=1/6, p(x1,Y = 2) = 3/6

p(x2,Y = 1) = 1/6, p(x2,Y = 2) = 1/6

$$MI(X,Y) = p(x1, Y = 1) * \log_2\left(\frac{p(x1, Y=1)}{p(x1) * p(Y=1)}\right) + p(x1, Y = 2) * \log_2\left(\frac{p(x1, Y=2)}{p(x1) * p(Y=2)}\right) + p(x2, Y = 1) * \log_2\left(\frac{p(x2, Y=1)}{p(x2) * p(Y=1)}\right) + p(x2, Y = 2) * \log_2\left(\frac{p(x2, Y=2)}{p(x2) * p(Y=2)}\right) = -1.87$$

- Evaluates the **information in one variable** that is **also present in another variable**
- **Marginal probabilities**
- **Joint probabilities**
- **Ranges** from **0** (independent) to $+\infty$
- **Symmetric**
  - MI(X, Y) = MI(Y, X) = MI(-X, Y) = MI(X, -Y)
  - Better suited to capturing nonlinear relationships between X and Y because it is based on the use of the logarithmic function

Table 2: Mutual information

	Domain known (x_1)	**Domain unknown (x_2)**	**marginal p**
Spam (y_1)	joint $p(x_1,y_1)$	joint $p(x_2,y_1)$	$p(y_1)$
No spam (y_2)	joint $p(x_1,y_2)$	joint $p(x_2,y_2)$	$p(y_2)$
marginal p	$p(x_1)$	$p(x_2)$	

Source of the table: Müller-Kett, 2023.

FEATURE PERMUTATION

1. **Select** a feature candidate
2. **Shuffle** the values in this column
3. **Train** a machine learning model
4. **Assess** the model's performance
5. Select the **next feature** candidate and repeat steps 2 to 4
6. The feature with the **largest model performance drop** is the **most informative**

Table 3: Feature permutation

X1	X2	X3	Y
…	…	…	…
…	…	…	…
…	…	…	…

Source of the table: Müller-Kett, 2023.

FEATURE VARIANCE

1. **Calculate the variances** of each feature
2. **Drop features** with **less** variance than a pre-defined **threshold**

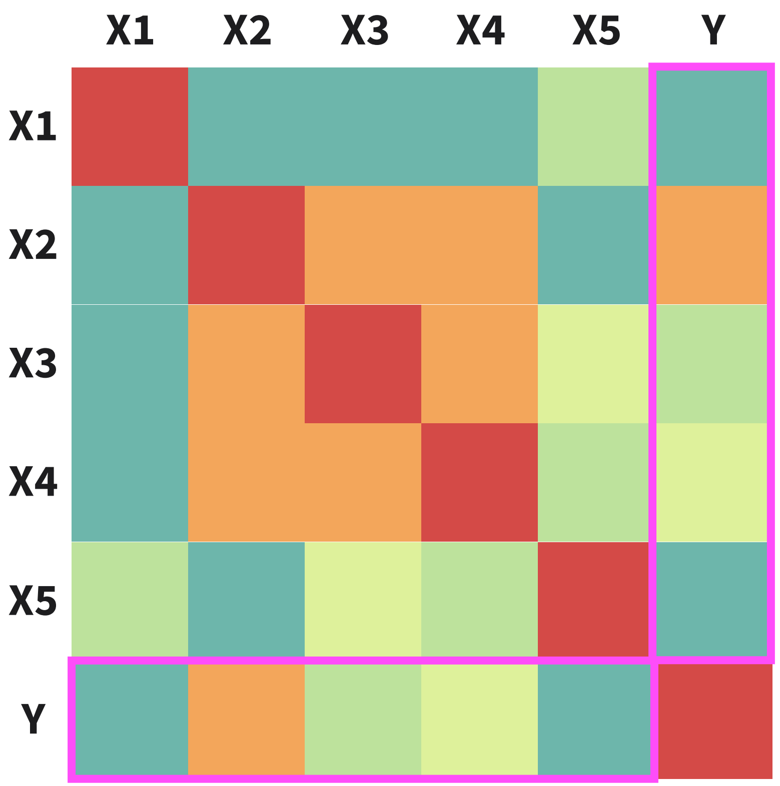
- Very **simple** and **fast** technique
- Useful to **quickly clean** a dataset for **informative columns**, e.g., remove singularity columns

Image 3: Singularity

Source of the image: Microsoft archive.

- **Covariance**
  - Simultaneous deviations of two variables from their mean
- **Pearson's correlation**
  - Normalization by the product of standard deviations
  - Range between -1 and 1
  - Assumes normal variable distributions
- **Spearman's correlation**
  - Based on value ranks
  - No normal distribution assumption

Image 4: Correlation matrix



Source of the table: Müller-Kett, 2023.

- Covariance measures a linear relationship between two variables

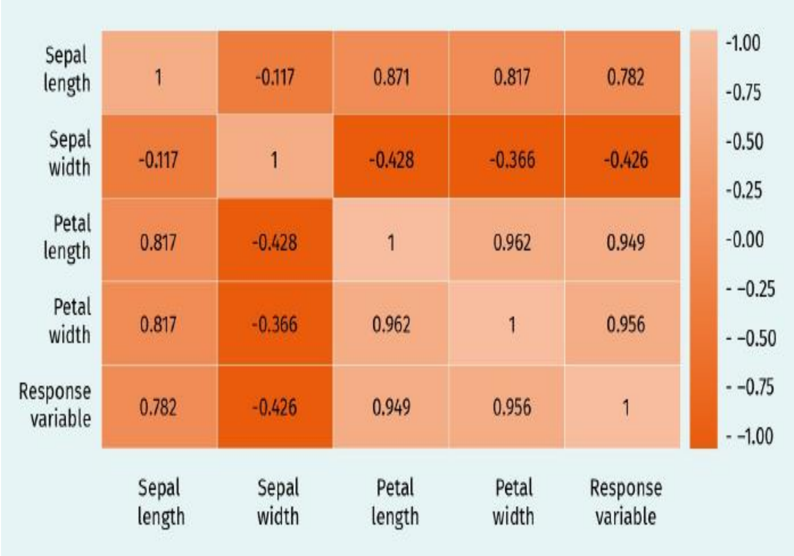
$$Cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \text{mean}_x)\cdot(y_i - \text{mean}_y)$$

where n is the number of rows in X and Y.

- Pearson's correlation
  - Dividing the covariance by the product of their standard deviations

$$\text{Pearson's correlation score}\Big(X,Y\Big) = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

  - Assumption: the data points in each column follow a Gaussian or normal probability distribution and the relationships are linear between the different columns

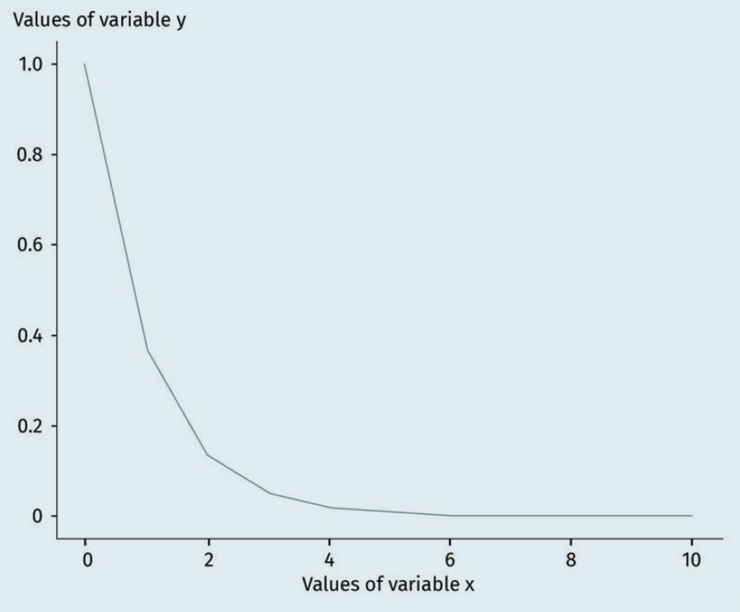


- Spearman's correlation

$$Spearman's\ correlation\ score\left(X,Y\right)=\frac{Cov(rank(X),rank(Y))}{\sigma_{rank(X)}\cdot\sigma_{rank(Y)}}$$

  - The rank of the values of a variable is obtained by sorting these values in decreasing order → requires a monotonic relationship between the two variables in the sense that when x1 > x2 then y1 < y2.

Nonlinear Correlation between Two Variables Represented as an Exponential Relationship



Pearson's and Spearman's Correlation for a Nonlinear Correlation Represented as an Exponential Relationship between Two Variables

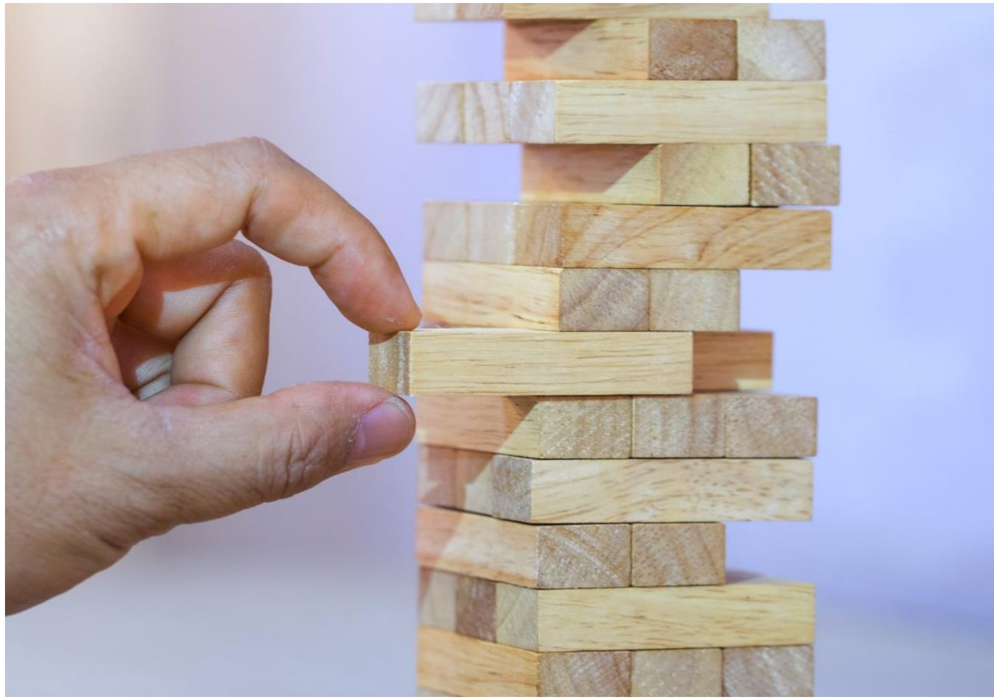
x	y	rank(x)	rank(y)
0	1	0	10
1	0.36	1	9
2	0.13	2	8
3	0.05	3	7
4	0.018	4	6
5	0.006	5	5
6	0.002	6	4
7	0.0009	7	3
8	0.0003	8	2
9	0.0001	9	1
10	0.0000	10	0
Pearson's correlation coefficient		Spearman's correlation coefficient	
–0.69		–1	

Source of the images: Müller-Kett, 2023.

RECURSIVE FEATURE SELECTION

1. Start with **all features (top-down)** or only **one feature (bottom-up)**
2. **Train** a machine learning model
3. **Assess** the model's performance
4. **Remove** one feature **(top-down)** or **add** one feature **(bottom-up)**
5. **Repeat** steps 2 and 3
6. Continue **until specified number of features** or **defined model performance** is reached

Image 5: Top-down recursive feature selection



Source of the image: Microsoft archive.

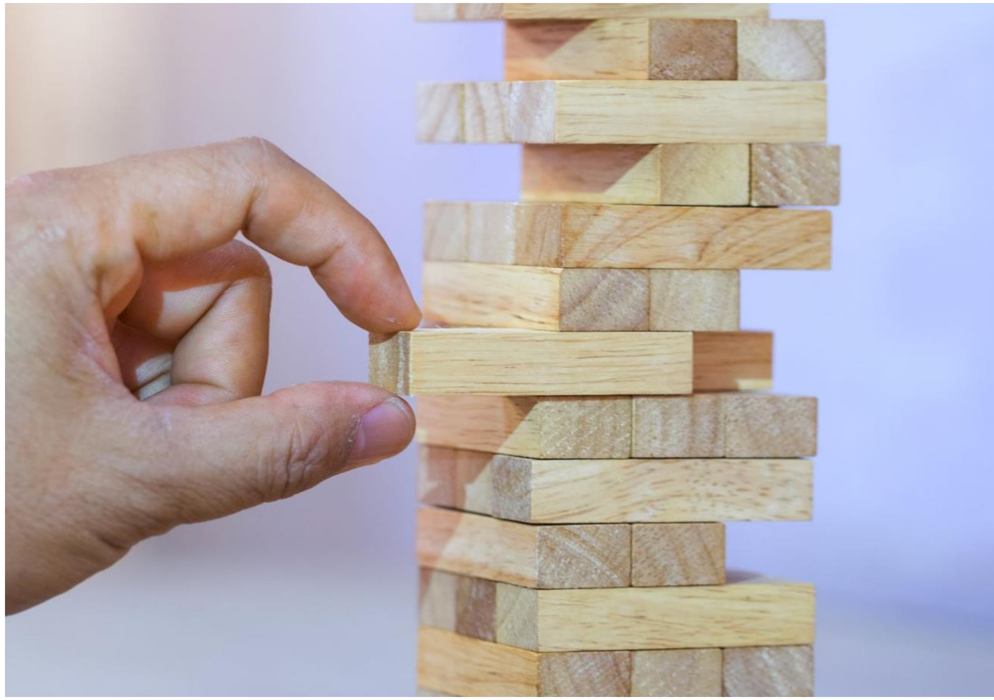
Exclusive Feature Selection (EFS)

- **Bottom-up** approach
- Best feature combination out of **all possible combinations of k features**
- Computationally **expensive**

Recursive Feature Elimination (RFE)

- **Top-down** approach
- Starts with **all possible feature combinations**
- Iterative **removes the least informative combination**
- Computationally **most expensive**

Image 5: Top-down recursive feature selection



Source of the image: Microsoft archive.

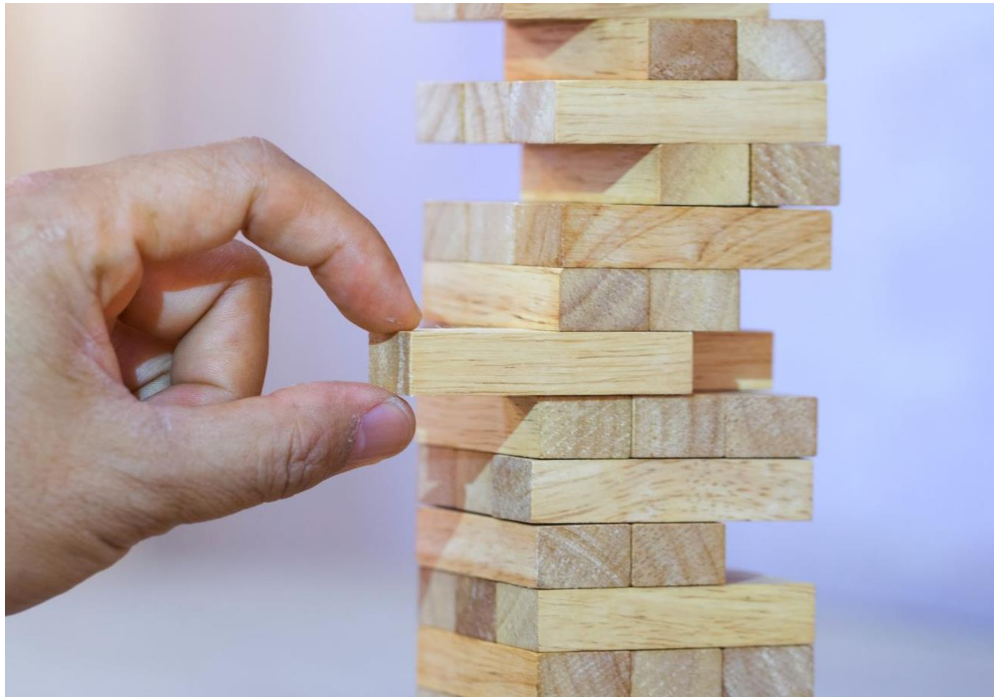
Sequential Forward Feature Selection (SFS)

- **Bottom-up** approach
- Calculates **singular feature importance** first
- Only **combines most important** features

Sequential Backward Feature Selection (SBS)

- **Top-down** approach
- Calculates feature importance for **combinations of all but one features**
- Sequentially **removes the feature not occurring** in the most important feature combination

Image 5: Top-down recursive feature selection



Source of the image: Microsoft archive.

REVIEW STUDY GOALS

Describe the **different techniques** used to **select relevant features**.

Explain how to **rank features** according to certain relevant **evaluation criteria**.

Select the best features in order to **maximize the performances** and **avoid overfitting** of the learned model.

SESSION 5

TRANSFER TASK

A start-up that **sells sustainable products in smaller stores worldwide** has been very successful in recent years.

To better understand **what makes satisfied customers**, a **survey** was conducted.

As a Data Scientist, you and your team are tasked to **determine the most relevant influences** that contribute to customer satisfaction.

Here, you can see the **first couple of rows** (there are **1000 in total**) and **columns** (there are **500 in total**) of the dataset provided to you:

Table. 4: Transfer task data sample

Has placed an order	Has received discount	Number of orders	Satisfied
Yes	No	2	Yes
Yes	Yes	1	No
Yes	Yes	7	Yes

- Discuss **potential problems** with this dataset that will be used to **train a machine learning model** and **solution strategies**. By doing so, also evaluate potential problems of **individual columns**.
- Marketing likes to provide a metric for the **effectiveness of giving discounts**, and they also want you to provide a **test statistic** for this metric. Describe which methods are suitable in this context.
- In addition to providing feature importance for each column, management asks you to provide **metrics for feature combinations**. This is a high-priority project, and you are equipped with the **most giant machines** available. Discuss which methods are appropriate for this use case.

Source of the table: Müller-Kett, 2023.

TRANSFER TASK
PRESENTATION OF THE RESULTS

Please present your results.
The results will be discussed in plenary.

1. Which one of the following characteristics applies to feature variance?
   a) wrapper feature selection technique
   b) multi-variate feature selection technique
   c) supervised feature selection technique
   d) unsupervised feature selection technique

2. Which one of the following characteristics applies to a correlation matrix?
   a) unsupervised and supervised feature selection technique
   b) supervised feature selection technique
   c) unsupervised feature selection technique
   d) wrapper feature selection technique

3. Which one of the following characteristics applies to recursive feature elimination?
   a) filter feature selection technique
   b) wrapper feature selection technique
   c) uni-variate feature selection technique
   d) unsupervised feature selection technique

LIST OF SOURCES

Text

Bejani, M., Gharavian, D., & Charkari, N. M. (2014). Audiovisual emotion recognition using ANOVA feature selection method and multi-classifier neural networks. *Neural Computing and Applications*, *24*(2), 399—412.

Images

Müller-Kett, 2021.
Müller-Kett, 2023.
Microsoft Archive.

Table

Müller-Kett, 2023.