A better decision tree for articles
Task 1

Add the correct articles to the following sentences or mark the absence of articles by −!

1. ___true positive rate was higher in ___method \( X \) than ___method \( Y \).

2. ___method \( X \) had ___higher true positive rate than ___method \( Y \).

3. ___memory means ___power or ___process of recalling.

4. \( X \) is ___algorithm which solves ___Travelling Salesman problem. ___algorithm \( X \) is ___fastest among all ___known \( TSP \) algorithms.

5. ___data set \( X \) follows ___Normal distribution with ___parameters \( \mu \) and \( \sigma^2 \). ___parameter \( \mu \) is ___mean of ___set \( X \) and ___parameter \( \sigma^2 \) is ___variance of ___\( X \).

6. ___problem \( X \) belongs to ___class \( P \), if it has ___polynomial time algorithm \( Y \). ___time complexity of ___algorithm \( Y \) is \( O(p(n)) \) where \( n \) is ___size of input and \( p \) is ___polynomial function.
7. In next section we introduce theory of Bloom filters.

8. To assess students’ program codes, we construct bug library. bug library contains all errors which have occurred in students’ programs.

9. Infinite time Turing machines extend idea of traditional Turing machines.

10. In pattern extraction we produce set of new attributes from original ones. goal is to find such set of attributes which describes data best. goodness of representation depends on modelling purpose, and in practice we have to define appropriate goodness measure.

11. In clustering analysis we divide data points into clusters such that all data points in one cluster are similar to each other but different from data points in other clusters.
12. ___episode is ___set of ___events which occur together. If ___order of ___events is fixed, ___episod is called serial.

13. There is always ___danger that ___model overfits. ___danger that ___model overfits is unavoidable.

14. ___main parts of ___computer are ___central unit, ___hard disk, and ___i/o devices. ___central unit is responsible for all ___computation.
Task 2

Are the following words countable or uncountable? Which articles can you use with them? Give example sentences!

- space
- requirement
- model
- program
- computation
- power
- capacity
- data
- information
- knowledge
- recognition
- software
- hardware
• code
• value
• property
• strength
• weakness
• use
• usability
Pronouns

Two important rules when you use pronouns:

1. When a pronoun refers to a noun in the preceding sentence, make sure that the referred is obvious!

2. Each pronoun should agree with the referant in number and gender.
Unclear references

- The simple pronouns – it, they, this, that, these, those – do often create ambiguities.

- Goal: the reader should not have to scan the previous sentence to understand what you mean.

- Recommendation: Avoid them, when possible! If you use them, always check twice that the meaning is not ambiguous!

- Never use ”those” – it usually a sign that the sentence is foggy.

 ”There was no difference in the accuracy of models between those which belonged to group A and those which belonged to group B. → ”The models in groups A and B were equally accurate.”
• Do not use ”it” to begin a sentence, if it is not absolutely clear, what it refers! (Exception: expressions like ”It is difficult to estimate...”)

• Hint: often you can replace ”this/these” + noun by ”the” + noun!
”This experiment demonstrated...” → ”The experiment demonstrated...”
Phrases

one – the other (singular)
some – the others (plural)

each other, e.g. ”X and Y affect each other”
This kind of + singular noun, e.g. ”This kind of system...”
If you want plural you have to say ”Systems of this kind...”

on one’s own, e.g. ”The students solved the task on their own”.

”All but one point belong to cluster 1”
”First of all, we have to initialize the parameters”
”On the one hand, the system is stable, on the other hand, it has poor accuracy”
”The initialization phase is time demanding. Otherwise the program is very efficient.”
Relative pronouns

Relative pronouns (who, which, that)
→ Section Relative clauses.
Adjectives

These seem to be well mastered, just two notes:

1. **Avoid vague adjectives!**

2. How to derive and use comparative and superlative forms?

Vague adjectives

- Do not use vague adjectives. Especially the adjectives which describe amounts (large, small, huge) are very context-sensitive!

- E.g. for statisticians, a data set of 500 rows is quite large, while for a data miner it is extremely small → numbers are more exact!

- The expressions become even vaguer, when you add modifiers ”quite”, ”rather”, ”very”, etc. Skip them always when possible!
Comparative and superlative

Basic rule: use -er/-est for short adjectives, and more/most for longer ones.

→ Read the exact rules from your material!
When you compare things

When you use the comparative, make clear what you are referring!

”Problem X is easier to solve” (than what?)

Basic structure:

\[ X \text{ is as efficient as } Y \ (X \text{ and } Y \text{ are equally efficient}) \]
\[ X \text{ is more efficient than } Y \]

Exceptional expressions:

\[ X \text{ is different from } Y \]
\[ X \text{ is similar to } Y \]
\[ X \text{ is the same as } Y \]
\[ X \text{ is inferior/superior to } Y \]
Verbs, nouns, pronouns, numerals, and adjectives compose the skeleton of sentences. The additional stuff consists of

- adverbs,
- prepositions, and
- conjunctions.
Adverbs


They express

- time (immediately, now, soon, later, next)
- place (here, there, everywhere)
- manner (easily, temporarily, well, poorly)
- degree (very, quite, ...) → Avoid in scientific texts!
- frequency (often, seldom, usually, sometimes)
- speaker’s attitude ”Fortunately, the data set is small, and function $f$ can be computed in real time.” → use sparsely!
Notes:

- **Recommendation**: Use expressive verbs and nouns which express the most of message, and as few adverbs/prepositional phrases as possible!

- Use introductory adverbs like ”fortunately, similarly, conversely, certainly” carefully, as a synonym to expressions ”it is fortunate” or ”in a similar manner”. Drop them if they are not needed.

- Notice that ”importantly” and ”interestingly” are not proper adverbs.
The position of adverbs in a sentence

The adverb can be

1. in the beginning, when you express time or attitude. E.g. ”Evidently, the students’ learning outcomes depend on their effort”, ”Later, we realized that…”

2. in the end, when you express way, time or place. E.g. “This problem occurs frequently in sparse data.”

3. in the middle, when you express frequency or attitude. Notice that already behaves in the same way. E.g. ”In knowledge discovery, we assume that the features have been already extracted”

An adverb should clearly refer to the word it modifies!
Parallel structures

Parallel structures are used to present parallel ideas.

**Parallel structure** = words, phrases, clauses or sentences combined by commas and/or conjunctions. Here we call the combined items as *parallel items*.

- Parallel items are combined by parallel conjunctions (and, or, but, ...).
- Notice that lists are also parallel structures!
• Often the parallel structure lists alternatives or makes some kind of comparison: the items belong to the same or similar classes or to two opposite classes.

• E.g.
"Method X has several advantages: it is easy to implement, it works in polynomial time, and it can use both numeric and categorical data."
contains two parallel structures: three advantages ("it is, it works, it can") in a list and "both numeric and categorical data"
Basic rules

The parallel structure should be consistent in two ways

- **Semantically**: the concepts referred by parallel items should be comparable, i.e. the comparison should make sense.

- **Syntactically**: the items should have similar grammatic structure. All of them should be either nouns, noun phrases, verb phrases, or clauses. In addition, they should be in the same form, e.g. you cannot combine ”to” + verb and a verb without ”to”.

  ”The problem is both hard to define and solve”

  → ”The problem is both hard to define and to solve”
Parallel items combined by conjunctions and, or, but

The most common form of parallel structures!

”The method has low space but high time requirement”
→ ”The method has low space requirement but high time requirement.

”The students were told to make themselves comfortable, to read the instructions, and that they should ask about anything they did not understand”
→ ”The students were told to make themselves comfortable, to read the instructions, and to ask about anything they did not understand”

”The results show that $X$ did not affect the error rate and the model overfitted the data”
→ ”The results show that $X$ did not affect the error rate and that the model overfitted the data”
Lists

Notice that elements in a list should be in a parallel form!

Example 1

”Boud [Bou02] has listed general characteristics which are typical for problem-based courses:

• Acknowledgement of learners’ experience.
• Emphasis on students taking responsibility of their own learning.
• Crossing of boundaries between disciplines.
• Focus on the processes of knowledge acquisition rather than the products of such processes.
• Change in staff role from instructor to facilitator.
• Students’ self- and peer assessment of learning.
• Focus on communication and interpersonal skills.”
Example 2

"The clustering methods can be divided into three categories:

1. *Hierarchical methods* construct a hierarchy of (typically) nested clusters.

2. *Partitioning methods* try to find optimal partitioning into a specified number of clusters.

3. *Probabilistic model-based clustering* tries to find the underlying probabilistic model which has produced the data."
Example 3

"The whole procedure is following:

1. Determine the number of clusters \( k \)
2. Choose parametric models (density functions \( f_j \)) for each of the clusters.
3. Determine the component probabilities \( \pi_k \) and parameters \( \theta_k \) from data.
4. Assign each point to the most probable cluster."
Example 4

If possible, avoid the following kind of list!

“According to O’Shea [OSh98], an intelligent tutoring system should be

• robust,
• helpfull
• simple,
• transparent
• flexible
• ...
• sensitive, and
• powerfull.”
Parallel items combined by conjunction pairs

Sometimes the parallel expression consists of two conjunctions like

- *between*...*and*,
- *both*...*and*,
- *either*...*or*,
- *neither*...*nor*, and
- *not only*...*but*.

The first conjunction should be immediately before the first part of the parallelism.
between – and

”between 20-22 years of age” → ”between 20 and 22 years of age”

”We recorded the difference between the students who completed the first task and the second task”
→ ”We recorded the difference between the students who completed the first task and the students who completed the second task.”

both – and

”The task is both easy to solve and efficient.”
(Doesn’t make any sense!)
→ The task is both easy to solve and can be solved efficiently.”

Or another structure:
”The task is easy and the solution is efficient.”
either – or

”The students either gave the worst answer or the best answer.”
→ ”The students either gave the worst answer or gave the best answer.” or ”The students gave either the worst answer or the best answer.”

neither – nor

In negative clauses → less often needed in sci-wri! (Say things in a positive way, when possible.)

”X solves the problems of traditional clustering algorithms. Neither outliers nor missing values affect the clustering quality.”

(Grammatically correct, but better to say: ”X solves the problems of traditional clustering algorithms. It is not sensitive to outliers or missing values.”)
not only – but (also)

”The task is not only easy to solve but also efficient”
→ ”The task is not only easy to solve but the solution is also efficient” or
”The task is not only easy to solve but it can also be solved efficiently”

Once again: say in a positive way, when possible – clearer!
On the one hand – on the other hand

- A special expression: can combine either clauses or parallel sentences!
- An affective way to describe opposite points, like advantages and disadvantages!

"On the one hand, a complex model can describe the data well, but on the other hand, it overfits easily."

"There is always a wrestling between the descriptive power and the generalization ability. On the one hand, too complex a model describes the data well, but it does not generalize to any new data. On the other hand, too simple a model generalizes well, but it does not describe the essential features in the data."
Parallel sentences

Numerating properties or ideas is an efficient way to create logical structures into paragraphs.

"$X$ model has three important properties: **First**, the model structure is easy to understand. This is a critical feature in adaptive learning environments, as we have noted before. **Second**, the model can be learnt efficiently from data. There are feasible algorithms for both numeric and categorial data. **Third**, the model tolerates noise and missing values."