- What is a **confusion matrix**? Name two **performance metrics** and describe how they can be calculated from the confusion matrix.
- Given 2 binary fingerprints we were asked to calculate the **tanimoto coeff**. And the **Dice coeff**. And explain what the variables are in each equation
- Calculate the Tanimoto coeff. For 2 non binary vectors/fingerprints
- How can **"Activity Landscape**" (AL) be defined? Name two different ways to represent AL.
- Why are similarity search methods capable of identifying structurally diverse compounds having similar activity?
- You're given 10 ranks with their corresponding "True or False" and you're asked to draw a **ROC curve** and calculate the hit rate.
- **Tversky coefficient** … we were asked for the formula**(single parameter variation**), and to define each variable in it, and what would make the **Tversky and Tanimoto** coefficient the same thing.
- Briefly compare **k-means and knn**.
- There are 2 ways for generating a chemical reference space .. **coordinate based** and **coordinate free** … we were asked to briefly explain both of them
- How and why is **"kernel trick**" used
- Given an **InCh**I string, we were asked to draw the molecule
- Given a **smiles** string, we were asked to draw the molecule
- There was a question about **cross validation** and how is it used to tune parameters (you can get back to the data science slides cause i think it was never mentioned in chemoinformatics :D)
- Manually execute **Ullmann algorithm** using the feasibility matrix method
- Describe how we update ranks in **CANGEN** algorithm (only the updating part)
- Given a big molecular structure, we were asked to order the first 9 atoms (The question asked us to write the beginning of the canonical smiles containing the first 9 atoms and number them accordingly) → no idea how to do it :DD
- This question I left blank so I really don't remember the exact wording but it was about "the paradigm of low chemical space dimensionality" .. and these were the questions Linus wrote on the whatsapp group ..they were referring to the same question but I cannot approve of it, cause i only read them once in the exam :D.

  "Name 2 purposes of lower dim. representations.
  Why can lower dimensional chemical representation clustering not perfectly represent chemical space.
  Name a profound unsupervised clustering Method, utilizing this.
  Name problems with correlation (in features / in lower dimensions)
  "
- How would u implement a numerical **discontinuity score** of compound datasets?