Guideline on how to prepare for answering questions at the oral exam

Questions at the oral exam are generally broad and more proposing to develop a theme than requiring a short (eg “yes/no”) answer.

Have a mindset of lecturing, as if you were presenting the topic to an aware but non-expert audience:

- From the topic, start by defining clearly what the problem is. Contextualize and give limits to the problem, maybe including an example in a sentence.
- You may give examples of the data that would serve the processing you will describe (see below). Mention any possible generalization (eg, “words in a document, but not only. Could be regions in an image”).
- Define clearly the terms you use. Provide the matching between their intuition and formal modelling. For example, provide the intuition between TF-IDF as a balance between representation and discrimination *and* the equation for the actual computation, including frequencies (counting)
- Be rigorous in the developments, write equations, not floating pieces of math. Make sure that you can answer questions on the details of whatever you write or present.
- Make associations with similar techniques or techniques using the same abstractions (eg low-rank models in LSI and recommendation systems). Or highlight differences with techniques solving the same problem with a different approach (eg various retrieval models).
- Use support examples to convey the intuition you want to present. Mention counter-examples when relevant.
- Relate to (identify) abstract mathematical tools. Eg: in LSI we model the linear decomposition of terms and documents into concepts using SVD (which is a generic abstract mathematical tool). (eg, LSI *is not* SVD, LSI *uses* SVD, LSI *boils down to* SVD on the term-document matrix).
- Use diagrams or equations whenever better than handwaving.
- You can also be *operational*: imagine you receive some data (that you describe). How would you go about processing this data to solve the problem? and why would you perform each step? (eg, how would you center the data, and why it is nice -or necessary- to center the data).
- Do not mention programs but rather mostly processes. You may mention well-known programs or libraries performing these operations (off-the-shelf or less high level)

Note: this guideline can be taken as a general way to present a notion to someone interested.