Questions :

🡪 Part 1 :

* Centered ? Nope Faire coeff X features ? Nope
* Pourquoi score carré ? -> pour variance
* PCA sur train ? 🡪 ouiii
* Which subset of data will you use to estimate your PC ? How will you apply it to unseen data?
* Normalization ? Avant, sur train. Si feature pas normalisé, plus de weight pour celui la. Feature entre 0 et 1 : variation de 0.9 enorme. Mais si feature entre 0 et 300, variation de 0.9 change rien.
* Que tirer des images ?

🡪 Part 2

* Keepout ?

🡪 Part 3 :

* Comment sortir Nrank et Npca optimal ? Sens de la boucle juste ?
* Regarder deuxième question

Previous guidesheet 4

🡪 Cross validation normal : diminution de param diminue overfitting ? Oui: moins de param = moins de samples et donc moins d’overfitting

🡪 Nested : compare median error of the hyperparam : optimal validation error or test-out error ? 🡪 Median supprime outliers