- Training using all backgrounds **weighted by cross section**
- Background mass is assigned randomly according to the **distribution of signal masses**
- Feature scaling done by **median/quartile** rather than mean/std
- Don’t understand how they **weight the signal**
- Exclusion plots from **score distributions**
- **Hyperparameter optimization** on 2000 random realizations
- **5-fold cross validation** splitting events in even/odd → remove high variance realizations
- Different **network architectures** for each channel (for us: ggF/VBF, merged/resolved)
- **ROC** as performance metric
- **Nesterov momentum** (what is it?)
- **Gaussian Process Regression** to pick hyperparameters (what is it?)
- They observe no effect of **dropout regularisation**
- **Overtraining plots** (what are they?)
- PNN response in terms of **Asimov significance** (what is it?) → gives nice plots (mass resolution, similar to what Gabriele is doing…)
- Asimov significance in the **mass/nomass comparison** (Martino’s studies)
- PNN **improves previous analyses** only below 450 GeV, and they claim this is because the previous analysis was sub-optimal, not because the PNN is better. However, the PNN interpolates
- Using **loss-vs-epoch plots** to assess under/over-training
- They speak about **generalisation error and generalisation gap** (to my understanding, the space between training and validation in the loss-epoch plot) and how this gives information. To be understood