5.1 ROC curve and AUC

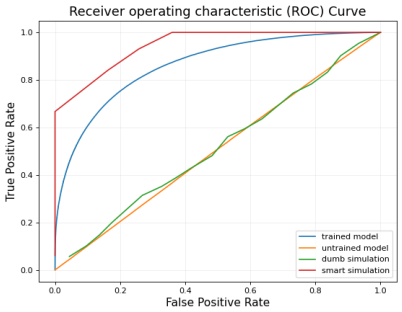


Figure 1: Receiver Operating Characteristic (ROC) curve for the dumb and smart simulation, as well as for the untrained and trained model on the basset test set. The trained model learned with 5 epochs and patience set to 2, batch size of 64 and a learning rate set to 0.002. We can notice that the dumb simulation describes the same as an untrained model. A straight line on the plot means that the prediction make is equivalent to a model which makes random prediction. The smart simulation curve looks more like the trained model one.

	AUC score
dumb simulation	0.50
smart simulation	0.94
untrained model	0.51
trained model	0.86

Table 1: Area Under Curve (AUC) for simulations and models on the basset test set. The AUC provides a way to quantify how well a model performs (i.e. having a high true positive rate and low false positive rate). AUC score is between 0 and 1 and correspond to the area under the curve as its name suggests. We can notice that for the dumb simulation has an AUC of 0.5 as well as the untrained model which as an AUC 0.51 which means that their performance is quite similar. Both the smart simulation (0.94) and the trained model (0.86) have a better AUC than the dumb simulation and untrained model respectively. That can be interpreted as the trained model is much more performant than the untrained one, as we could expect.

5.2 ROC Column-wise normalized PWM & 5.3 Maximum activation filters

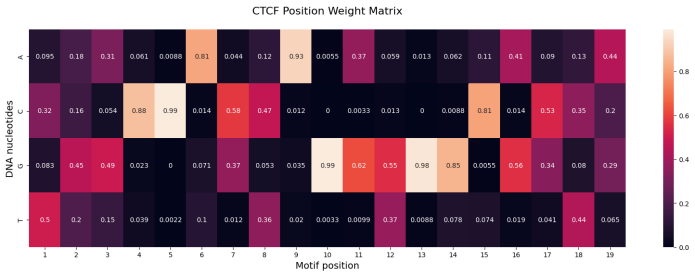


Figure 2: Column-wise normalized Position Weight Matrix for the CTF Motif.

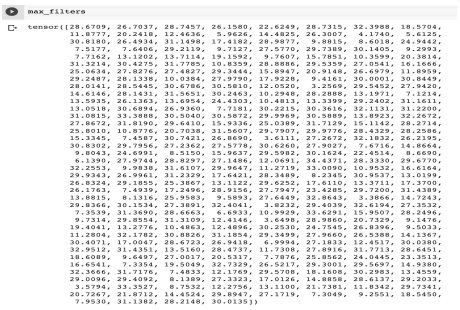
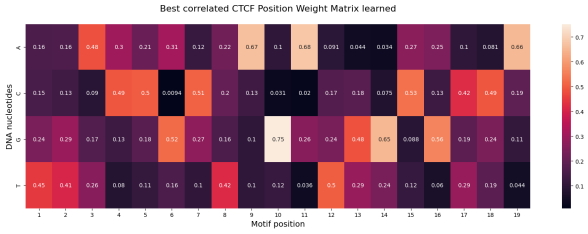
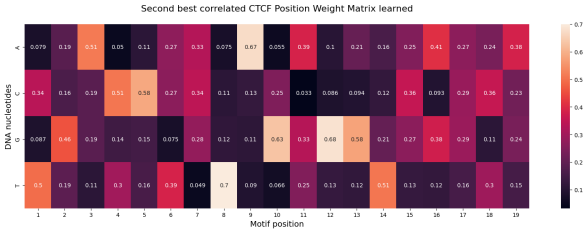


Figure 3: Maximum values found that activate both of the 300 filter of the first convolution on the basset model across all examples from basset test set.

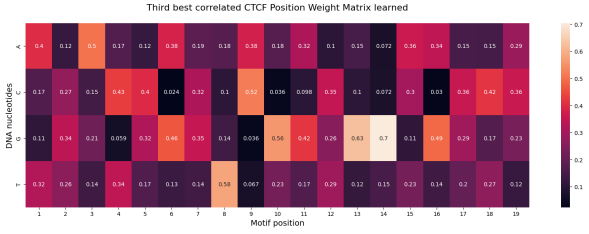
5.4 & 5.5



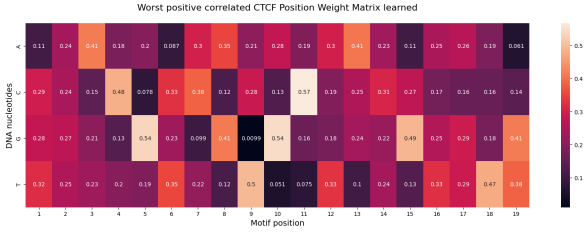
(a) Pearson correlation coefficient: 0.75



(b) Pearson correlation coefficient: 0.70



(c) Pearson correlation coefficient: 0.67



(d) Pearson correlation coefficient: 0.0006

Figure 4: Top 3 and most uncorrelated Column-Wise normalized Position Weight Matrices (PWM) of the CTCF motif based on the 300 filters learned from the first convolution layer of the basset model. Each of the PWM were compared to the original Fig. 2 (extracted and column-wise normalized from MA0139.1.jaspar file) using Pearson's coefficient correlation. The four PWM are represented by heatmaps. To build those heatmaps, the 300 vector maximum activation of basset's first convolution layer across all the test set was used to build the PWM by summing all the patches convoluted of the test set that were upper to the activation threshold (the maximum activation divided by 2 for given filter). Finally, after column-wise normalizing that resulting 300 PWM, we can extract the top 3 positive correlated PWM (a) (b) (c) and the worst one (d) used for comparison.

rem: The model that made those heatmaps was trained with same config as Fig.**??** but with a learning rate set to **0.01**