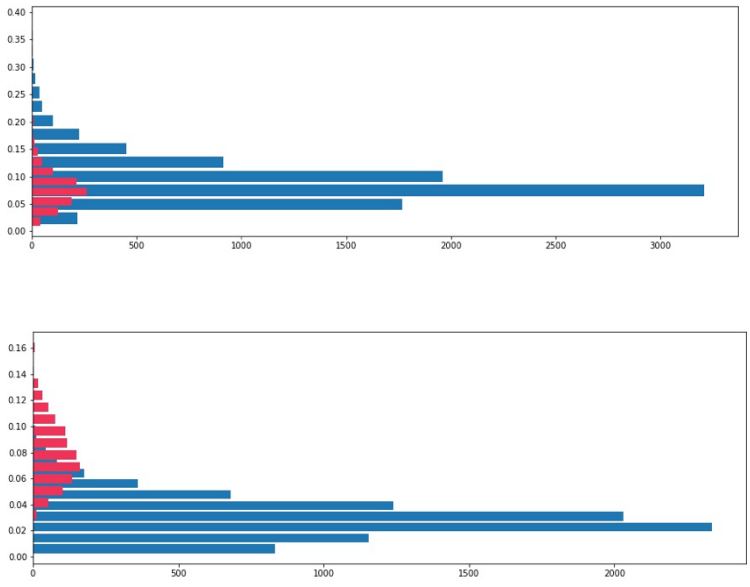
L’exercice consiste à mettre dix fois de côté un des dix digits du jeu de données MNIST et à entraîner un modèle de détection d’anomalie sur le jeu de données déplété du chiffre, de manière à détecter celui-ci comme une anomalie lors de la phase de test sur le jeu de données contenant la totalité des chiffres.

Rappelons que le jeu de donnés de train contient 60 000 images labellisées de 1 à 10 et le jeu de données test en contient 10 000. L’exercice impose de re labelliser les images correspondant aux 9 chiffres gardés avec le label 1 pour normal et le chiffre enlevé avec le label 0 correspondant à l’anomalie.

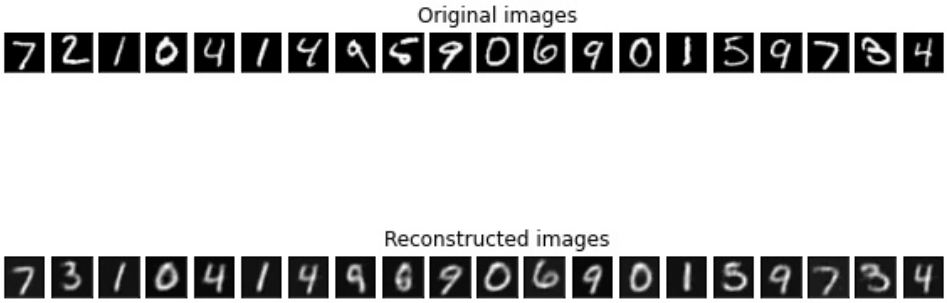
Ici le premier modèle testé a été un AutoEncoder (AE) classique mais c’est finalement un AutoEncoder variationnel (VAE) qui a servi de modèle pour l’exercice. De par son architecture l’AE constitue un modèle favorable aux problématiques de détection d’anomalie en partant du postulat que, si il est assez bien entraîné, alors l’erreur MSE calculée sur l’ensemble des pixels de l’image reconstituée en sortie de l’AE et ceux de l’image originale sera minime, permettant ainsi d’établir un seuil d’anomalie qui correspond à une valeur de MSE au dessus de laquelle l’image est « anormale ». En effet on peut aisément imaginer que l’AutoEncoder se trompera d’avantage dans la reconstitution d’une image qu’il n’a « jamais vue».

Dans le cas présent le premier modèle type AE n’a cependant pas permis d’obtenir des reconstitutions assez fidèles pour avoir des range de MSE significativement différents entre les images normales et anormales, rendant impossible la détermination d’une valeur de MSE comme seuil d’anomalie.

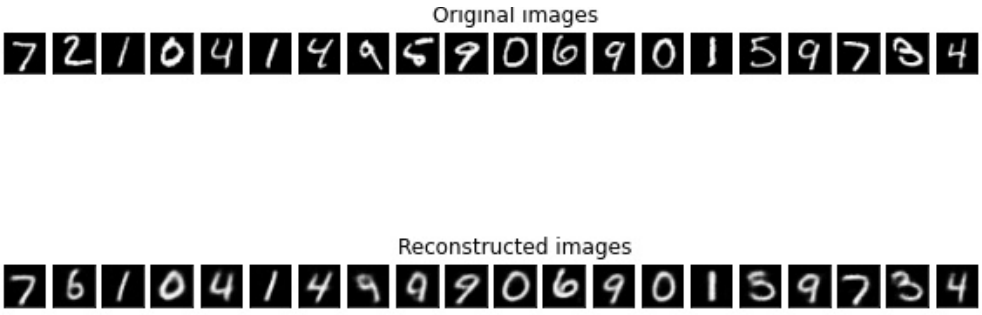
En revanche un VAE a permis de d’avantage minimiser les erreurs de reconstitution permettant de décaler la distribution des erreurs MSE des images normales par rapport à celles obtenues avec les images anormales. Ci-dessous est représenté une exemple des répartitions des images dans les valeurs de MSE selon le label du fichier (bleu : normal et rose : anormal) permettant de comparer la différence des distributions selon si elles ont été obtenues avec l’AE (en haut) ou le VAE (en bas) lorsque le chiffre 2 est l’anomalie :



Fidélité des reconstitutions obtenues avec l'AE (2 anormal) :



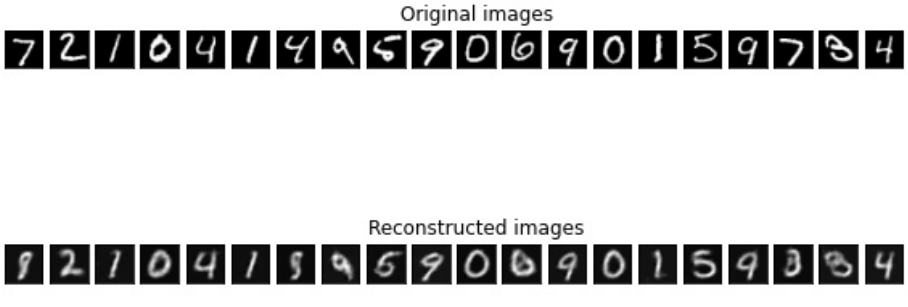
Fidélité des reconstitutions obtenues avec le VAE (2 anormal) :



En effet on voit bien sur ces reconstitutions des 20 1eres images du jeu test (toujours dans le contexte du chiffre 2 qui est anormal) que ces dernières sont plus fidèles avec le VAE qu'avec l'AE.

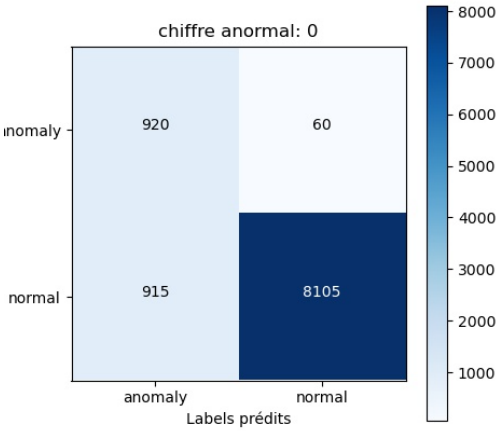
On observe aussi que certaines images sont fondamentalement plus dures à reconnaître pour le modèle dans le cas de chiffres « mal écrits »:
Ici par exemple la 9ème image, qui correspond à un 5, est mal reconstituée avec les deux modèles et fait partie des images qui « tirent » la MSE vers le haut. Le VAE visiblement limite le nombre d'images concernées par ce problème en comparaison à l'AE. L'exemple ci-dessous l'illustre très bien, avec le chiffre 7 qui est anormal : on voit bien que l'AE a du mal à reconstituer d'autres chiffres pourtant normaux contrairement au VAE :

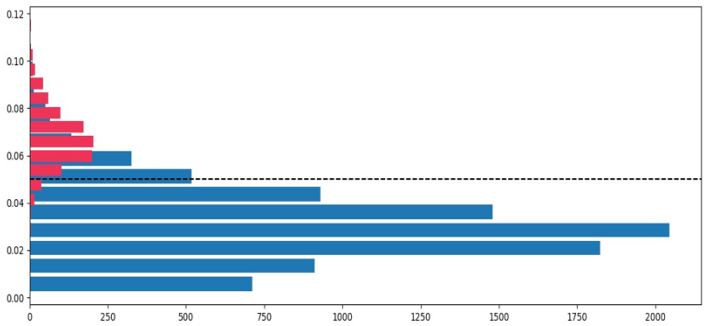
Fidélité des reconstitutions obtenues avec l'AE (chiffre 7 anormal) :



Fidélité des reconstitutions obtenues avec le VAE (chiffre 7 anormal) :

Ce sont les chiffres anormaux permettant d'établir le seuil le plus discriminant en termes de répartitions des MSE entre normal et anormal qui donnent bien évidemment les meilleurs score AUC, comme ici le zéro qui lorsqu'il est l'anomalie, permet d'obtenir un score AUC de 92% en prenant un seuil d'anomalie égal à 0.05 (ligne pointillée sur les histogrammes) :





Un cas particulier est le 1 lorsqu'il est l'anomalie : étant donné qu'il est très « facile » à reconstituer les modèles ne font finalement pas plus d'erreurs dans la reconstitution d'un 1 même s'ils n'ont pas été entraînés avec auparavant que lorsqu'ils doivent reconstituer des chiffres très ambigus. Par conséquent pour discriminer l'anomalie, il a fallu à l'inverse, décider que c'est en dessous d'un seuil d'erreur de mse que le chiffre est un 1 étant donné la facilité de sa représentation. Les chiffres 4 et 9 sont associés à des MSE beaucoup trop similaires entre normal et anormal rendant la détermination d'un seuil quasiment impossible, et les performances qui en résultent tirent l'AUC moyenne vers le bas.

Finalement les chiffres anormaux permettant les meilleurs séparations des distributions de MSE et donc les meilleurs performances avec le VAE sont : 0, 2, 5, 6 et 8.
A l'inverse l'autre moitié des chiffres : 1, 3, 4, 7 et 9 donnent des résultas médiocres voire similaires à ceux obtenus avec une classification random dans le cas du 9.
Les chiffres qui peuvent être confondus entre eux si ils sont mal écrits sont, dans le cas des anormaux donnant un bon score :
-le 0 avec le 6 ; le 6 avec le 8 ; le 6 avec le 5 ; et le 2 avec le 5

Pour expliquer pourquoi environ la moitié des chiffres donnent d'excellents scores on peut imaginer que la proportion de chiffres « mal écrits » ambigus est plus importante dans ces catégories des 0, 2, 5, 6 et 8 : par exemple le 0 et le 6 donnent d'excellents scores sans doute parce qu'il ya une très importante proportion d'images labellisées 0 qui ressemblent à des 6 et vice versa. Ainsi, enlever le zéro par exemple, non seulement va entraîner une diminution du nombre d'images contribuant à faire monter la MSE, mais en plus va permettre au modèle d'être « plus sûr de lui » pour reconstituer les 6. A l'inverse on peut imaginer qu'enlever le 4 ou le 9 pourrait tirer la MSE vers le haut parce que les 4 et les 9 sont la majorité du temps très bien écrits, ainsi les confusions entre les 2 seraient initialement assez rares.