

# GEI898 – Apprentissage profond avancé

## Devoir 1: Détection d'anomalies

François Grondin

Département de génie électrique et génie informatique

Faculté de génie

Université de Sherbrooke

Janvier 2026

# Détection d'anomalies

## Ensemble de données

Nous utiliserons l'ensemble de données Shuttle qui représente des mesures de neuf capteurs. Chaque mesure est associée à une classe, qui représente un état sur la navette. La classe 1 représente ce qu'on considèrera comme un état sain, et les autres classes des états fautifs.

Capteurs									Classe
1	2	3	4	5	6	7	8	9	
53	0	82	0	52	-5	29	30	2	1
37	0	76	0	28	18	40	48	8	1
37	0	79	0	34	-26	43	46	2	1
85	0	88	-4	6	1	3	83	80	5

Tableau 1: Exemples de mesures dans l'ensemble de données

## Entraînement

La tâche consiste à entraîner un auto-encodeur sur **les valeurs saines seulement** du sous-ensemble d'entraînement.

Entraînement avec le sous-ensemble d'entraînement de données saines

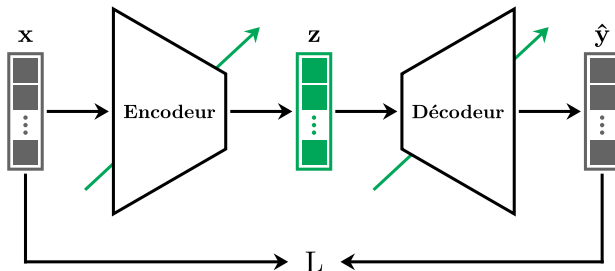


Figure 1: Procédure d'entraînement de l'auto-encodeur

## Test

Calcul du coût avec les données saines et fautives du sous-ensemble de validation.

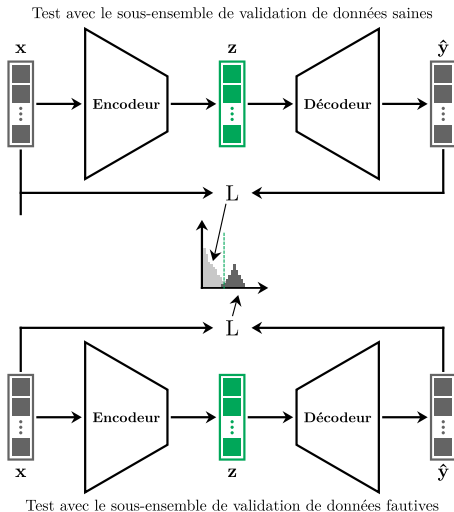


Figure 2: Procédure de test avec l'auto-encodeur

## Performances

Les performances sont mesurées en terme F-mesure Ce score est obtenu avec l'équation suivante:

$$F_1 = \frac{2VP}{2VP + FP + FN}, \quad (1)$$

où VP est le nombre de vrais positifs, FP le nombre de faux positifs et FN le nombre de faux négatifs.

En modifiant la valeur du seuil  $T_0$ , il est possible de classifier un échantillon comme positif ( $L \geq T_0$ ) ou négatif ( $L < T_0$ ).

Dans le cadre de ce projet, il est normalement possible d'atteindre une F-mesure supérieure à 90%.

## Astuce

Parmi les astuces à employer dans le cadre de ce devoir, il est important de considérer une normalisation des données. Pour chaque capteur, il est possible de normaliser les données à partir de la moyenne et la déviation standard du sous-ensemble d'entraînement.

$$x'_{c,i} = \frac{x_{c,i} - \mu_c}{\sigma_c}, \quad \mu_c = \frac{1}{I} \sum_{i=1}^I x_{c,i}, \quad \sigma_c = \sqrt{\frac{1}{I} \sum_{i=1}^I (x_{c,i} - \mu_c)^2} \quad (2)$$

Ceci permet une meilleure stabilité numérique lors de l'entraînement, et également une fonction de coût mieux équilibrée pour chaque capteur.

# Livrables

## Présentation

Une présentation orale avec support visuel qui comprend entre autres:

- ▶ L'architecture du réseau de neurones;
- ▶ La stratégie d'entraînement du réseau;
- ▶ L'approche pour trouver le seuil optimal;
- ▶ Une courbe de F-mesure en fonction du seuil.

## Période de questions

L'équipe professorale posera diverses questions à l'oral sur des thématiques en lien avec le devoir pour valider votre compréhension des concepts.