UNIVERSITE DE YAOUNDE I
FACULTE DES SCIENCES
DEPARTEMENT D'INFORMATIQUE

INF4127 : OPTIMISATION II

RAPPORT
TPE N° 1

Analyse des fonctions de pertes

Encadrant :	**Pr. Paulin MELATAGIA**	
Etudiants :	ETOUNDI TSANGA Elihu Frédéric	22Y567
	MVOGO MONDOMAN Franck Stéphane	20Y058
	ABANDA Armand Wilfried	21T2487

YAOUNDE, OCTOBER 2025

Contents

1 Introduction **4**

2 Gradient et Convexité des fonctions de pertes **4**

2.1 Expressions des gradients 4

2.1.1 Erreur Quadratique Moyenne (MSE) 4

2.1.2 Entropie Croisée Binaire (BCE) 4

2.1.3 Entropie Croisée Catégorielle (CCE) 5

2.1.4 Perte de Huber 5

2.2 2. Propriétés de convexité 5

2.2.1 Erreur Quadratique Moyenne 6

2.2.2 Entropie Croisée Binaire 6

2.2.3 Entropie Croisée Catégorielle 6

2.2.4 Perte de Huber 6

3 Cas pratique **6**

1 Introduction

Une fonction de perte est un processus mathématique qui quantifie la marge d'erreur entre la prédiction d'un modèle et la valeur cible réelle. Elle est un moyen mesurable d'évaluer les performances et la précision d'un modèle d'apprentissage automatique.

Dans cet exercice, nous analysons quatre fonctions de perte : l'Erreur Quadratique Moyenne (MSE), l'Entropie Croisée Binaire (BCE), l'Entropie Croisée Catégorielle (CCE) et la Perte de Huber. Pour chacune, nous calculons les gradients, étudions la convexité, les appliquer sur des jeux de données pour la régression et la classification, représentons les courbes des fonctions, et déterminons l'équation de la tangente à une courbe de niveau.

2 Gradient et Convexité des fonctions de pertes

2.1 Expressions des gradients

Nous supposons que $\hat{y}$ est la prédiction du modèle et y la valeur réelle. Les gradients sont calculés par rapport à $\hat{y}$.

2.1.1 Erreur Quadratique Moyenne (MSE)

La fonction de perte est définie par :

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

Pour une observation unique, $L_{MSE} = (\hat{y} - y)^2$.

Le gradient est :

$$\frac{\partial L_{MSE}}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}(\hat{y} - y)^2 = 2(\hat{y} - y)$$

Gradient : $\nabla L_{MSE} = 2(\hat{y} - y)$.

2.1.2 Entropie Croisée Binaire (BCE)

La fonction de perte est :

$$L_{BCE} = -\left[y\log(\hat{y}) + (1-y)\log(1-\hat{y})\right]$$

où $\hat{y} \in (0,1)$ et $y \in \{0,1\}$.

Le gradient est :

$$\frac{\partial L_{BCE}}{\partial \hat{y}} = -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}$$

Gradient : $\nabla L_{BCE} = -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}$.

2.1.3 Entropie Croisée Catégorielle (CCE)

Pour K classes, la perte est :

$$L_{CCE} = -\sum_{k=1}^{K} y_k \log(\hat{y}_k)$$

où $y_k = 1$ si la classe k est la vraie classe, 0 sinon.

Le gradient pour une classe k est :

$$\frac{\partial L_{CCE}}{\partial \hat{y}_k} = -\frac{y_k}{\hat{y}_k}$$

Gradient : $\nabla L_{CCE} = -\frac{y_k}{\hat{y}_k}$.

2.1.4 Perte de Huber

La perte de Huber est définie par :

$$L_{Huber} = \begin{cases} \frac{1}{2}(\hat{y}-y)^2 & \text{si } |\hat{y}-y| \leq \delta \\ \delta|\hat{y}-y| - \frac{1}{2}\delta^2 & \text{sinon} \end{cases}$$

où $\delta > 0$.

Le gradient est :

$$\nabla L_{Huber} = \begin{cases} \hat{y}-y & \text{si } |\hat{y}-y| \leq \delta \\ \delta \cdot \text{sign}(\hat{y}-y) & \text{sinon} \end{cases}$$

2.2 2. Propriétés de convexité

Une fonction est convexe si sa Hessienne est semi-définie positive.

2.2.1 Erreur Quadratique Moyenne

$$L_{MSE} = (\hat{y} - y)^2$$
$$\frac{\partial^2 L_{MSE}}{\partial \hat{y}^2} = 2 > 0$$

La fonction est strictement convexe.

2.2.2 Entropie Croisée Binaire

$$\frac{\partial^2 L_{BCE}}{\partial \hat{y}^2} = \frac{y}{\hat{y}^2} + \frac{1-y}{(1-\hat{y})^2}$$

Puisque $\hat{y}, 1-\hat{y} \in (0,1)$ et $y \in \{0,1\}$, la dérivée seconde est positive. La fonction est convexe.

2.2.3 Entropie Croisée Catégorielle

$$\frac{\partial^2 L_{CCE}}{\partial \hat{y}_k^2} = \frac{y_k}{\hat{y}_k^2} \geq 0$$

Les dérivées croisées sont nulles, donc la Hessienne est semi-définie positive. La fonction est convexe.

2.2.4 Perte de Huber

Pour $|\hat{y} - y| \leq \delta$, $\frac{\partial^2 L_{Huber}}{\partial \hat{y}^2} = 1 > 0$. Pour $|\hat{y} - y| > \delta$, la dérivée seconde est 0. La fonction est convexe, mais non strictement convexe.

3 Cas pratique

Le notebook est disponible ici