TP Série 1 : Analyse du Gradient et de la Convexité des Fonctions de Perte

UE : INF4127 - Optimisation II

Membres du groupe
KUITCHE AROLLE NACHARD (22T2931)
KONZOU SODEA ALAN PEREC (22T2957)
KENFACK LEKANE FRANK (22T2841)
VOUKENG DJIOKENG CHRISTIAN ROUSSEL (22U2053)

UNIVERSITÉ DE YAOUNDÉ 1
Faculté des Sciences, Département d'Informatique

2 octobre 2025

Objectif de l'Exercice

- **Calcul symbolique** des gradients des fonctions de perte classiques.
- Étude des **propriétés de convexité** pour chaque fonction.

Importance

La convexité est cruciale en optimisation pour garantir que la descente de gradient converge vers un minimum global.

Notations Utilisées

- y : **Vraies étiquettes** (vecteur cible).
- ŷ : **Prédictions** du modèle (vecteur).
- N : Nombre d'échantillons.
- $e_i = y_i - \hat{y}_i$: Erreur pour l'échantillon i.
- δ : Paramètre de la **Perte de Huber**.
- $\nabla_{\hat{y}} L$: Gradient par rapport aux prédictions ŷ.

1. Erreur Quadratique Moyenne (MSE)

- **Formule :**

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

→ **Gradient (Vectoriel) :**

$$\nabla_{\hat{\mathrm{y}}} L_{MSE} = \frac{2}{N}(\hat{\mathrm{y}} - \mathrm{y})$$

→ **Convexité :** La seconde dérivée est $\frac{\partial^2 L_{MSE}}{\partial \hat{y}_i^2} = \frac{2}{N}$.

Conclusion

La MSE est une fonction **strictement convexe** ($\frac{2}{N} > 0$).

2. Entropie Croisée Binaire (BCE)

- **Formule :**

$$L_{BCE}(\mathrm{y}, \hat{\mathrm{y}}) = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

→ **Gradient (par composante) :**

$$\frac{\partial L_{BCE}}{\partial \hat{y}_i} = \frac{1}{N} \left(\frac{\hat{y}_i - y_i}{\hat{y}_i(1 - \hat{y}_i)} \right)$$

→ **Convexité :** La seconde dérivée est $\frac{\partial^2 L_{BCE}}{\partial \hat{y}_i^2} = \frac{1}{N} \left[\frac{y_i}{\hat{y}_i^2} + \frac{1 - y_i}{(1 - \hat{y}_i)^2} \right]$.

Conclusion

La BCE est **strictement convexe** (car la seconde dérivée est toujours positive pour $\hat{y}_i \in (0, 1)$).

3. Entropie Croisée Catégorielle (CCE)

- **Formule :**

$$L_{CCE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik}\log(\hat{y}_{ik})$$

→ **Gradient (par composante $\hat{y}_{ik}$) :**

$$\frac{\partial L_{CCE}}{\partial \hat{y}_{ik}} = -\frac{1}{N}\frac{y_{ik}}{\hat{y}_{ik}}$$

→ **Convexité :** La seconde dérivée est $\frac{\partial^2 L_{CCE}}{\partial \hat{y}_{ik}^2} = \frac{1}{N}\frac{y_{ik}}{\hat{y}_{ik}^2}$.

Conclusion

La CCE est **convexe** (≥ 0).

Note : La convexité est généralement étudiée par rapport aux logits (entrées du Softmax) pour l'optimisation.

4. Perte de Huber

- **Formule (pour un échantillon i, avec $e_i = y_i - \hat{y}_i$) :**

$$L_\delta(e_i) = \begin{cases} \frac{1}{2} e_i^2 & \text{si } |e_i| \leq \delta \\ \delta(|e_i| - \frac{1}{2}\delta) & \text{si } |e_i| > \delta \end{cases}$$

$\rightarrow$ **Gradient (par composante $\hat{y}_i$) :**

$$\frac{\partial L_{Huber}}{\partial \hat{y}_i} = -\frac{1}{N} \cdot \begin{cases} y_i - \hat{y}_i & \text{si } |e_i| \leq \delta \\ \delta \cdot \text{sgn}(y_i - \hat{y}_i) & \text{si } |e_i| > \delta \end{cases}$$

$\rightarrow$ **Convexité :** La seconde dérivée est positive ou nulle : $\frac{\partial^2 L_{Huber}}{\partial \hat{y}_i^2} = \frac{1}{N}$ (si $|e_i| < \delta$) ou 0 (si $|e_i| > \delta$).

Conclusion

La Perte de Huber est **convexe**. Elle est moins sensible aux outliers que la MSE tout en conservant la convexité.

Conclusion : Synthèse des Résultats

Résumé

Toutes les fonctions de perte étudiées sont **convexes** par rapport aux prédictions $\hat{y}$, ce qui est fondamental pour les algorithmes d'optimisation comme la descente de gradient.

Fonction de Perte	**Gradient**	**Propriété de Convexité**
MSE	$\frac{2}{N}(\hat{y} - y)$	Strictement Convexe
BCE	$\frac{1}{N}\left(\frac{\hat{y}-y}{\hat{y}\odot(1-\hat{y})}\right)$	Strictement Convexe
CCE	$-\frac{1}{N}\frac{y}{\hat{y}}$ (Élément par élément)	Convexe
Huber	Expression par cas	Convexe

Merci pour votre attention !