

Compte Rendu TP3

Atelier ML



Prepared by

MOUNA BEN ZAIED

ONS OUHIBI

RT 4 G 2

Question 1

-Compléter les parties manquantes :

```
# Standardiser Les valeurs des pixels
X_train_std, params = standardize(X_train, columns=range(X_train.shape[1]), return_params=True)
X_test_std = standardize(X_test, columns=range(X_test.shape[1]), params=params)

# Initialiser Le réseau de neurones
nn1 = MLP(eta=0.1, epochs=300, hidden_layers=[50], n_classes=None, momentum=0.0,
          l1=0.0, l2=0.0, dropout=1.0, decrease_const=0.0, minibatches=1, random_seed=None, print_cost=True)

# Apprendre Les caractéristiques tout en imprimant La progression
nn1.fit(X_train_std, y_train)

# Afficher La courbe d'apprentissage
plt.plot(range(len(nn1.cost_)), nn1.cost_)
plt.ylabel('Cost')
plt.xlabel('Epoch')
plt.show()

# Afficher La précision sur Les ensembles d'entraînement et de test
print('Train Accuracy: %.2f%%' % (100 * nn1.score(X_train_std, y_train)))
print('Test Accuracy: %.2f%%' % (100 * nn1.score(X_test_std, y_test)))
```

Question 2

-Expliquer le fonctionnement du modèle:

- Le modèle de réseau de neurones multi-couches (MLP) appliqué à la classification des chiffres manuscrits dans la base de données MNIST est conçu pour apprendre des représentations complexes à partir des images en entrée et les utiliser pour prédire de manière précise le chiffre correspondant à chaque image. L'architecture du modèle comprend une couche d'entrée représentant les pixels de l'image, une ou plusieurs couches cachées qui apprennent des caractéristiques hiérarchiques, et une couche de sortie avec des neurones correspondant aux différentes classes de chiffres (0 à 9).

Au cours de l'entraînement, le modèle procède à l'ajustement de ses poids par le biais de la descente de gradient stochastique, où le taux d'apprentissage (η) influence l'amplitude des ajustements. L'application de régularisation L1 et L2 vise à prévenir le surajustement en restreignant la complexité du modèle. L'utilisation du momentum accélère l'apprentissage en tenant compte des tendances antérieures des mises à jour de poids. Le dropout, une méthode de régularisation, intervient pour éviter la coadaptation des neurones. Les mini-lots facilitent l'optimisation, tandis que le mélange des données avant chaque époque prévient la formation de cycles indésirables dans le processus d'apprentissage.

L'entraînement du modèle est illustré par la courbe d'apprentissage, affichant comment le coût diminue au fil des époques. Les performances sont évaluées sur des ensembles d'entraînement et de test, mesurant la capacité du modèle à généraliser sur des données non vues. Ces caractéristiques combinées font du MLP un puissant outil de classification pour les tâches complexes comme la reconnaissance de chiffres manuscrits.

Question 3

- Manipuler les paramètres, comparer la performance du modèle et interpréter les résultats:

Comparaison des performances en fonction des changements de paramètres :

1. **Eta:** Une valeur plus faible (0.001) a amélioré la performance par rapport à 0.1 et 0.01, ce qui peut indiquer que des pas plus petits dans l'espace des poids conduisent à une convergence plus stable.
→ Un taux d'apprentissage plus faible engendre une convergence plus lente mais aboutit à une meilleure stabilité et précision finale.
Trouver le juste équilibre est essentiel, car des valeurs trop élevées peuvent conduire à une convergence rapide, mais au détriment de la précision du modèle.
2. **Epochs :** Une augmentation du nombre d'epochs a conduit à de meilleures performances (300 à 600), mais il peut y avoir un point où l'augmentation supplémentaire n'apporte plus d'amélioration significative.
Cependant, il est important de noter qu'il existe un point optimal où l'ajout d'epochs supplémentaires pourrait ne plus apporter d'amélioration significative. Cela peut résulter d'un risque de surajustement.

3.Hidden Layer : L'ajout d'une couche cachée avec 100 unités a conduit à des performances supérieures par rapport à une couche cachée de 25 ou 50 unités. Cette observation suggère que l'augmentation de la complexité du modèle, en ajoutant davantage d'unités à la couche cachée, a favorisé une meilleure capture des caractéristiques discriminantes dans les données d'entraînement.

4.L1 et L2 : Lorsque nous avons renforcé la régularisation avec des valeurs de 0.01 pour L1 et L2, nous avons constaté une amélioration des performances par rapport à des valeurs plus élevées comme 0.1. Cependant, il est important de noter que la performance n'a pas été considérablement meilleure par rapport à l'absence de régularisation (0.0). Cela suggère qu'une régularisation plus intense a été bénéfique pour éviter le surajustement, mais il y a un équilibre à trouver. Une régularisation excessive peut limiter la capacité du modèle à apprendre des motifs complexes, entraînant une baisse des performances.

5.Dropout : Un dropout de 1.0, indiquant aucune suppression, a montré de meilleures performances que 0.5. Cette observation peut être attribuée à divers facteurs tels que la complexité du modèle et la quantité de données disponibles. Il est important de noter que l'efficacité du dropout peut varier en fonction de la nature spécifique du problème et du modèle. Dans ce contexte particulier, l'absence de dropout a conduit à de meilleurs résultats.

6.Changements simultanés : Les meilleures performances ont été obtenues en combinant tous ces ajustements simultanément. Cela souligne l'efficacité de l'approche globale dans l'optimisation du modèle. Lorsque des modifications multiples sont appliquées de manière cohérente, elles peuvent se compléter mutuellement pour améliorer la capacité du modèle à apprendre des représentations significatives et à généraliser à de nouvelles données.

Interpretation:

L'ajustement simultané de tous les paramètres a abouti à la meilleure performance, soulignant l'importance de considérer l'interaction globale entre les paramètres pour optimiser le modèle.

Question 4

-Appliquer différents modèles et comparer les résultats:

- Lors de l'évaluation de plusieurs modèles pour la tâche de classification binaire (classe 0 vs. autres classes) sur le jeu de données MNIST, deux modèles ont été comparés : le MLP (MultiLayer Perceptron) et la régression logistique. Le MLP, configuré avec un taux d'apprentissage (η) de 0.001, 300 epochs et une couche cachée de 100 unités, a affiché une précision de 92.58% sur l'ensemble d'entraînement et de 92.42% sur l'ensemble de test. En parallèle, la régression logistique, avec un taux d'apprentissage (η) de 0.1 et 50 epochs, a montré des performances nettement supérieures, avec une précision de 99.80% sur l'ensemble d'entraînement et de 99.84% sur l'ensemble de test.
- Cette comparaison met en évidence la différence de complexité entre les deux modèles. Alors que le MLP offre une plus grande flexibilité avec une couche cachée, la régression logistique, malgré sa simplicité, montre des performances exceptionnelles dans cette tâche spécifique. Ces résultats soulignent l'importance de choisir un modèle adapté à la nature de la tâche et à la complexité des données.