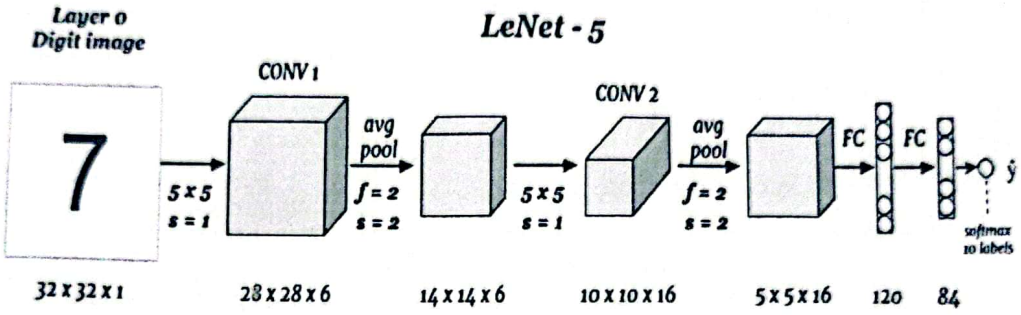
Université de Carthage
Ecole Supérieure de la Statistique et de l'Analyse de l'Information

Examen de Machine Learning ; 3 ème année du cycle de formation d'ingénieurs
Durée de l'épreuve : 1 heure 30 - Documents non autorisés
Nombre de pages : 2 - Date de l'épreuve : 6 janvier 2025

Exercice 1 : Déterminer le nombre de paramètres de l'architecture LeNet-5 représentée par la figure suivante :



Exercice 2 : Indiquer les deux principaux apports de l'architecture Inception de GoogleNet.

Exercice 3 : On considère la fonction suivante du package `stable_baselines3`. Commenter les paramètres de cette fonction.

```
DQN('MlpPolicy', "CartPole-v1", verbose=1, learning_starts=1000,
            target_update_interval=10,
            policy_kwargs={'net_arch': [256, 256]},
             tensorboard_log="./cartpole_tensorboard/")
```

Exercice 4 : QCM sur le Reinforcement Learning (RL)

Question 1 : Dans le RL, le composant qui représente l'environnement et définit la dynamique du système est appelé :

a) Policy
b) Reward
c) Transition Model
d) Value Function

Question 2 : Dans un problème de RL, que représente la fonction de récompense ?

a) Une estimation des transitions entre les états.
b) Une mesure immédiate du succès ou de l'échec d'une action effectuée.
c) Une politique optimale pour maximiser les gains.
d) Une approximation de la valeur future d'un état.

Question 3 : Dans le cadre de Q-Learning, l'équation de mise à jour des valeurs Q inclut :

a) Le gradient du modèle de l'environnement

b) Une approximation basée sur des heuristiques
c) Une récompense instantanée et une estimation future
d) Un calcul exact de la politique optimale

Question 4 : L'équation de Bellman sert principalement à :
a) Calculer une politique stochastique
b) Décomposer un problème en sous-problèmes récursifs
c) Maximiser directement la fonction de récompense
d) Décrire les algorithmes supervisés

Question 5 : Le dilemme **exploration vs exploitation** implique :
a) Trouver un équilibre entre explorer de nouvelles actions et exploiter les meilleures actions connues
b) Maximiser les récompenses futures tout en minimisant les coûts
c) Utiliser des modèles explicites pour simuler des environnements inconnus
d) Définir la meilleure fonction de valeur possible

Question 6 : Quel algorithme de RL n'a pas besoin d'un modèle de l'environnement ?
a) Value Iteration
b) Monte Carlo Tree Search
c) Q-Learning
d) Dynamic Programming .

Question 7 : Dans l'approche actor-critic, quel est le rôle du "critic" ?
a) Prendre des décisions sur les actions à effectuer.
b) Évaluer la performance de l'agent par rapport à la politique suivie.
c) Optimiser directement la politique d'action.
d) Réduire les erreurs de prédiction de l'environnement.

Question 8 : L'algorithme de Policy Gradient vise à :
a) Optimiser directement les paramètres de la politique
b) Évaluer la politique optimale à l'aide d'un modèle d'environnement
c) Maximiser les récompenses futures en ajustant les valeurs Q
d) Calculer une fonction de récompense pour les états terminaux

Question 9 : Un épisode dans RL se termine lorsque :
a) L'agent maximise la récompense instantanée
b) Un état terminal est atteint
c) L'agent ne reçoit plus de récompense
d) Une politique optimale est déterminée

Question 10 : L'utilisation d'une méthode comme le replay buffer dans le Deep Q-Learning permet :
a) De stocker les transitions pour une utilisation future et réduire la corrélation des données.
b) De calculer directement la politique optimale à chaque étape.
c) D'améliorer l'exploration en augmentant l'échantillonnage aléatoire.
d) D'augmenter les performances en augmentant le taux d'apprentissage.

Question 11 : Quel est l'objectif principal de la méthode Proximal Policy Optimization (PPO) dans le RL ?
a) Maximiser directement la fonction de récompense immédiate.
b) Stabiliser l'apprentissage en limitant les mises à jour drastiques de la politique.
c) Calculer une politique optimale en utilisant une équation de Bellman.
d) Minimiser l'erreur quadratique entre les valeurs Q prédites et observées.