

Optimisation et

Certificat Chef de Projet IA

Quizz - 2022/2023



Instructions

- Ce court questionnaire permet de valider le module Optimisation et recherche opérationnelle réalisé lors de la semaine du 10 au 14 octobre 2022.
- Vos réponses individuelles à ce questionnaire sont à transmettre par mail à clement.royer@lamsade.dauphine.fr.

Questions sur la partie 1 - Optimisation convexe

Question 1 Parmi les affirmations ci-dessous portant sur la programmation linéaire, laquelle est fausse ?

- a) Un programme linéaire est toujours convexe.
- b) Un programme linéaire est toujours réalisable.
- c) Il est possible de résoudre des programmes linéaires avec plusieurs millions de variables.
- d) Il est possible de résoudre des programmes linéaires avec des algorithmes de points intérieurs.

Question 2 Lorsque l'on cherche un modèle linéaire d'un jeu de données, il peut arriver qu'il n'existe pas de tel modèle qui colle parfaitement aux données. Dans ce cas, on cherche donc un modèle collant aux données le mieux possible en résolvant...

- a) un système linéaire;
- b) un programme quadratique non convexe;
- c) un système de KKT;
- d) un problème aux moindres carrés linéaires.

Questions sur la partie 2 - Optimisation non convexe

Question 3 Parmi les affirmations suivantes, laquelle est **fausse** ?

- a) Tout minimum local d'une fonction convexe est aussi un minimum global.
- b) Une fonction continue fortement convexe a un unique minimum global.
- c) Une fonction non convexe possède toujours des minima locaux non globaux.
- d) Un minima local est un point stationnaire d'ordre 2.

Question 4 Supposons que l'on applique une descente de gradient avec une longueur de pas bien choisie à un problème fortement convexe. Alors,...

- a) l'algorithme converge vers un point selle;
- b) l'algorithme converge en vitesse $\mathcal{O}(1/\sqrt{K})$;
- c) l'algorithme ne converge pas;
- d) l'algorithme converge vers un minimum global.

Questions sur la partie 3 - Gradient stochastique

Question 5 Pour entrainer un réseau de neurones on minimise généralement une somme finie de fonctions. Dans le cadre d'un problème de classification d'images, chaque terme de la somme représente:

- a) un pixel dans une image;
- b) une image du training set;
- c) toutes les images qui ont le même label;
- d) un batch de gradients.

Question 6 Pour minimiser une somme finie, on utilise généralement un algorithme de gradient stochastique car...

- a) la variance dans l'estimation du gradient permet de converger plus rapidement;
- b) c'est plus rapide de calculer le gradient d'un seul terme de la somme;
- c) c'est plus facile d'analyser un algorithme qui utilise des opérations aléatoires;
- d) tous les gradients sont à peu près les mêmes.

Questions sur la partie 4 - Optimisation sans(?) dérivées

Question 7 Parmi les approches suivantes, laquelle **n'est pas** équivalent au calcul d'un gradient lorsque la fonction est dérivable ?

- a) *Itération de gradient proximale;*
- b) *Différentiation symbolique;*
- c) *Calcul de sous-gradient;*
- d) *Backpropagation dans un réseau de neurones.*

Question 8 Les méthodes d'optimisation sans dérivées basées sur des modèles sont généralement plus performantes que des approches type recherche directe car...

- a) *elles n'évaluent pas la fonction à optimiser;*
- b) *elles réutilisent des évaluations déjà effectuées;*
- c) *elles reposent sur des modèles probabilistes;*
- d) *elles exploitent la statistique bayésienne.*

Questions sur la partie 5 - Optimisation à grande échelle

Question 9 L'algorithme *Adam*, qui est un des meilleurs algorithmes actuellement pour entraîner des réseaux de neurones, se caractérise par:

- a) *une descente par coordonnées stochastique;*
- b) *une combinaison de gradient stochastique avec momentum et un rescaling par coordonnées;*
- c) *un gradient stochastique sans momentum avec une procédure de rescaling pour diminuer la variance.*

Question 10 La méthode *ADMM* (*Alternating direction of multipliers*)...

- a) *permet d'optimiser certains problèmes en parallèle si on peut négliger les contraintes;*
- b) *est une méthode de descente par coordonnées stochastique;*
- c) *permet d'optimiser certains problèmes en parallèle lorsque le dual est séparable.*