

Optimisation et RO

Certificat Chef de Projet IA

Quizz - 2023/2024

Instructions

- Ce court questionnaire permet de valider le module Optimisation et recherche opérationnelle réalisé lors de la semaine du 9 au 13 octobre 2023.
- Chaque question ne possède qu'une réponse correcte.
- 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 *Quel est l'intérêt principal de formuler un problème de sciences des données comme un programme convexe (type LP, QP, SDP, CP) ?*

- a) *Il existe des solveurs numériques bien établis dédiés à ce type de problèmes.*
- b) *Il existe des algorithmes de points intérieurs avec garanties théoriques pour résoudre ce type de problèmes.*
- c) *On peut toujours trouver une solution au problème.*
- d) *On peut toujours résoudre le problème avec plusieurs millions de variables.*

Question 2 *On cherche à coller à un jeu de données via un modèle linéaire. Laquelle des affirmations suivantes est alors fausse ?*

- a) *Calculer le modèle par les moindres carrés est adapté au cas d'erreurs gaussiennes.*
- b) *On peut calculer l'approximation en norme ℓ_1 en résolvant un programme linéaire.*
- c) *On ne peut jamais calculer un modèle linéaire qui interpole parfaitement les données.*
- d) *Il existe toujours un modèle qui donne la meilleure approximation de Chebyshev.*

Questions sur la partie 2 - Optimisation non convexe

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

- a) Il existe des fonctions non convexes pour lesquelles tout minimum local est global.
- b) Il existe des fonctions convexes qui ont un unique minimum global.
- c) Il existe des fonctions convexes qui possèdent des maxima locaux non globaux.
- d) Il existe des fonctions non convexes qui possèdent des maxima locaux et des minima locaux.

Question 4 Supposons que l'on applique une descente de gradient avec une longueur de pas bien choisie à un problème convexe. Laquelle des affirmations ci-dessous est alors **fausse** ?

- a) L'algorithme converge en vitesse $\mathcal{O}(1/K)$.
- b) L'algorithme converge vers la valeur optimale du problème.
- c) L'algorithme peut être accéléré pour obtenir une vitesse de convergence optimale.
- d) L'algorithme n'est pas applicable à une fonction fortement convexe.

Questions sur la partie 3 - Gradient stochastique

Question 5 Parmi les propositions ci-dessous, laquelle **n'est pas une raison** pour utiliser les méthodes de gradient stochastique ?

- a) La taille du jeu de données peut être très grande.
- b) L'accès aux données peut être coûteux.
- c) Le gradient par rapport à une partie des données peut être une bonne direction.
- d) La descente de gradient ne converge pas sur ces problèmes.

Question 6 Parmi ces approches, laquelle est considérée comme la moins coûteuse dans un contexte de somme finie ?

- a) Appliquer une descente de gradient.
- b) Appliquer l'algorithme du gradient stochastique avec mini-batch.
- c) Appliquer l'algorithme du gradient accéléré.
- d) Appliquer l'algorithme du gradient stochastique de base.

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

Question 7 Parmi les approches suivantes, laquelle est équivalente au processus dit de backpropagation ?

- a) Différences finies + itération gradient proximal;
- b) Calcul du sous-différentiel complet + itération de sous-gradient;
- c) Différentiation automatique+ itération de gradient stochastique;
- d) Codage à la main des dérivées + itération de descente de gradient.

Question 8 Laquelle de ces affirmations sur l'optimisation sans dérivées est **fausse** ?

- a) Les algorithmes génétiques peuvent être performants sur des problèmes de petite dimension (2-3).
- b) Les méthodes basées sur des modèles exploitent plus l'information des valeurs de fonctions déjà obtenues que celles de recherche directe.
- c) Un algorithme d'optimisation sans dérivées ne peut utiliser aucune information sur la fonction, et la traite forcément comme une "boîte noire".
- d) Les dernières avancées en optimisation sans dérivées sont basées sur l'interaction avec d'autres domaines comme les approches de bandits.

Questions sur la partie 5 - Optimisation à grande échelle

Question 9 Quel est l'intérêt principal de la méthode SVRG ?

- a) Elle fonctionne en parallèle;
- b) Elle utilise le momentum;
- c) Elle est moins coûteuse que le gradient stochastique;
- d) Elle permet de réduire la variance du gradient stochastique.

Question 10 Dans quel contexte **n'est-il pas** pertinent d'utiliser une approche de descente par coordonnées ?

- a) Lorsque le problème est séparable;
- b) Lorsque le problème est partiellement séparable;
- c) Lorsque le problème est non lisse;
- d) Lorsque la dimension du problème est trop élevée pour stocker un gradient en mémoire.

Solutions

Question 1 La bonne réponse est a) : *Il existe des solveurs numériques bien établis dédiés à ce type de problèmes.* Même s'il existe des approches théoriques permettant de résoudre ces problèmes, l'intérêt principal de ces formulations est qu'elles peuvent être données telles quelles à des solveurs du marché.

Question 2 La bonne réponse est c) : *On ne peut jamais calculer un modèle linéaire qui interpole parfaitement les données.* Si le modèle a plus de paramètres qu'il y a d'exemples dans le jeu de données, il est en général possible d'interpoler les données.

Question 3 La bonne réponse est c) : *Il existe des fonctions convexes qui possèdent des maxima locaux non globaux.* Une fonction convexe ne possède par définition que des minima globaux, c'est-à-dire que tous ses points stationnaires (en lesquels le gradient est nul) correspondent forcément à des minima locaux, et non à des maxima locaux non globaux.¹

Question 4 La bonne réponse est d) : *L'algorithme n'est pas applicable à une fonction fortement convexe.* L'intérêt de l'algorithme de descente de gradient est précisément de pouvoir s'appliquer à tout type de fonction dérivable, que cette fonction soit convexe, non convexe ou fortement convexe.

Question 5 La bonne réponse est d) : *La descente de gradient ne converge pas sur ces problèmes.* L'intérêt du gradient stochastique réside dans sa capacité à converger vers une solution en utilisant moins d'accès aux données, mais la descente de gradient converge également.

Question 6 La bonne réponse est d) : *Appliquer l'algorithme du gradient stochastique de base.* Celui-ci ne requiert qu'un accès à un point du jeu de données à chaque itération.

Question 7 La bonne réponse est c) : *Différentiation automatique+ itération de gradient stochastique.* Lors du processus de backpropagation, le gradient de la fonction de perte évaluée en la fournée (*batch*) d'exemples utilisée est calculé par différentiation automatique.

Question 8 La bonne réponse est c) : *Un algorithme d'optimisation sans dérivées ne peut utiliser aucune information sur la fonction, et la traite forcément comme une "boîte noire".* Il est très fréquent en optimisation sans dérivées que l'on traite la fonction comme une "boîte grise", dont on connaît une partie de la structure que l'on peut alors utiliser dans un algorithme d'optimisation.

Question 9 La bonne réponse est d) : *Elle permet de réduire la variance du gradient stochastique.* L'acronyme SVRG signifie *Stochastic Variance Reduced Gradient*.

Question 10 La bonne réponse est c) : *Lorsque le problème est non lisse.* La descente par coordonnées repose en effet sur l'utilisation de coordonnées du gradient, ce qui implique donc que celui-ci existe.

¹On notera que, dans le cadre d'une fonction constante, tout minimum global est aussi un maximum global.