

TD 3 - Descente de gradient

Exercice 1 – Apéro

Q 1.1 Parmi les fonctions suivantes, lesquelles sont convexes :

$$f(x) = x \cos(x), g(x) = -\log(x) + x^2, h(x) = x\sqrt{x}, t(x) = -\log(x) - \log(10 - x) ?$$

Q 1.2 Soit une application linéaire $f \in \mathbb{R}^n \rightarrow \mathbb{R}$; rappeler ce qu'est le gradient de $f : \nabla f(\mathbf{x})$. Donner le gradient de $f(\mathbf{x}) = 2x_1 + x_2^2 + x_2x_3$

Q 1.3 Exprimer $\nabla_{\mathbf{x}}(f(\mathbf{x}) + g(\mathbf{x}))$, $\nabla_{\mathbf{x}}t f(\mathbf{x})$.

Donner l'expression de $\nabla_{\mathbf{x}}b' \mathbf{x}$ avec $b \in \mathbb{R}^d$ et $\nabla_{\mathbf{x}} \mathbf{x}' A \mathbf{x}$ pour A symétrique .

Exercice 2 – Régression linéaire

Soit un ensemble de données d'apprentissage $\mathcal{D} = \{\mathbf{x}^i, y^i\}_{i=1, \dots, N}$, $\mathbf{x}^i \in \mathbb{R}$, $y^i \in \mathbb{R}$.

Par convention que l'on suivra dans toute la suite du cours, la matrice de données sera notée X , où chaque ligne correspond à un exemple. La matrice Y des réponses est donc une matrice colonne ; la matrice W des poids également. L'erreur sur \mathcal{D} sera notée $C(W)$.

Q 2.1 Résolution analytique

Q 2.1.1 Rappeler le principe de la régression linéaire. Quelle fonction d'erreur $C(W)$ est utilisée ?

Q 2.1.2 Quelles sont les dimensions des matrices X , W et Y ? Rappeler la formulation matricielle de l'erreur.

Q 2.1.3 Trouver analytiquement la matrice W solution de la régression linéaire, qui minimise $C(W)$.

Q 2.1.4 Même question si l'on considère maintenant une machine linéaire avec biais. Quelle est la valeur optimale du biais w_0 dans ce cas ?

Q 2.2 Rappeler le principe de l'algorithme de descente du gradient. Donner son application au cas de la régression linéaire.

Q 2.3 On considère dans la suite un problème à 2 dimensions.

Q 2.3.1 Tracer l'espace des paramètres en 2D. Positionner arbitrairement les points \mathbf{w}^0 , point initial, et \mathbf{w}^* , solution analytique du problème. Etant donnée la nature quadratique du coût, tracer les iso-contours de la fonction de coût dans l'espace des paramètres. Quelle est la forme de la fonction de coût $C(\mathbf{w}^0)$ dans l'espace des paramètres ?

Q 2.3.2 Dessiner le vecteur $\nabla C(\mathbf{w}^0)$. A quoi correspond ce vecteur géométriquement ?

Exercice 3 – Régression logistique

Q 3.1 On considère un ensemble de données X muni d'étiquettes binaires $Y = \{0, 1\}$. En régression logistique, on considère que le log-rapport des probas conditionnelles $p(y|x)$ peut être modélisé par une application linéaire : $\log \left(\frac{p(y|x)}{(1-p(y|x))} \right) = \theta \cdot x$.

- Quel est le but ?
- Quelle étiquette prédire pour x si $\theta \cdot x > 0$?
- En déduire le type de frontière que la regression logistique permet de déterminer

- Que vaut $p(y|x)$? Tracer la fonction $p(y|x)$ en fonction de $\theta.x$.

Q 3.2 Pour une dimension x_i , quelle est l'influence de sa valeur pour $p(y|x)$? Quelle est la limite de la régression logistique ?

Q 3.3 Soit W les paramètres recherchés. Quelle est l'expression de la vraisemblance conditionnelle de W par rapport à un exemple (x, y) ? La log-vraisemblance? Et dans le cas d'un ensemble d'exemples \mathcal{D} ?

Q 3.4 Proposer un algorithme pour résoudre le problème de la régression logistique.

Exercice 4 – Optimisation d'un modèle gaussien par descente de gradient

Nous disposons ici d'un jeu de données non-étiquetées : $\mathcal{D} = \{\mathbf{x}_i\}_{i=1,\dots,N}$, $\mathbf{x}_i \in \mathbb{R}^d$.

Nous souhaitons apprendre en mode non supervisé un modèle gaussien correspondant aux données de \mathcal{D} . Le modèle gaussien est défini par un ensemble de paramètres $\{\mu, \Sigma\}$

Q 4.1 Exprimez la log-vraisemblance en supposant les exemples de \mathcal{D} statistiquement indépendants.

Q 4.2 Solution analytique

Q 4.2.1 Que vérifie la solution W^* du maximum de vraisemblance? Montrez que la solution W^* du maximum de vraisemblance correspond à la moyenne et la covariance empirique des données \mathcal{D} dans le cas où Σ est une matrice diagonale.

Q 4.3 Méthode de gradient

Q 4.3.1 Déterminez le gradient de la vraisemblance en un point W_0 .

Q 4.3.2 Ecrire deux algorithmes de gradient batch et stochastique permettant d'apprendre une loi gaussienne à partir de \mathcal{D} .