TD 2

Exercice 1 – Rappels sur la convexité

Q 1.1 Parmi les fonctions suivantes, lesquelles sont convexes :
$f(x) = xcos(x)$, $g(x) = -log(x) + x^2$, $h(x) = x\sqrt{x}$, $t(x) = -log(x) - log(10 - x)$?

Q 1.2 Soit une application linéaire $f \in \mathbb{R}^n \to \mathbb{R}$; rappeler ce qu'est le gradient de f : $\nabla f(\mathbf{x})$. Donner le gradient de $f(\mathbf{x}) = 2x_1 + x_2^2 + x_2 x_3$

Q 1.3 Exprimer $\nabla_{\mathbf{x}}(f(\mathbf{x}) + g(\mathbf{x}))$, $\nabla_{\mathbf{x}} tf(\mathbf{x})$.
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 $\mathcal{D} = \{\mathbf{x^i}, y^i\}_{i=1,\ldots,N}, \mathbf{x^i} \in \mathbb{R}^d, y^i \in \mathbb{R}$.
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 (5 points) – Regression tri-logistique

On s'attaque à un problème de classification à 3 classes en utilisant un modèle adapté de la régression logistique : $P(y = k|\mathbf{x}) = \frac{e^{\mathbf{w}_k^t \mathbf{x}}}{\sum_{j=1}^3 e^{\mathbf{w}_j^t \mathbf{x}}}$ pour $k = 1, 2, 3$.

Q 3.1 Quel rapport avec la régression logistique ?

Q 3.2 On veut estimer les paramètres par un critère de vraisemblance sur un ensemble de données $\{(\mathbf{x}^i, y^i) \in \mathbb{R}^d \times \{1, 2, 3\}\}_{i=1}^N$. Donnez le problème d'optimisation associé.

Q 3.3 Proposez un algorithme d'estimation pour les paramètres $\mathbf{w}_k$ et donnez le en pseudo-code.

Q 3.4 On remarque dans les expériences conduites une forte tendance au sur-apprentissage. Proposez une méthode pour contrôler le sur-apprentissage et le changement de formulation du critère d'estimation. Donner le nouvel algorithme.