

Remarque : la généralisation multidimensionnelle, aussi connue sous le nom de la méthode de Newton-Raphson, a la règle de mise à jour suivante :

$$\theta \leftarrow \theta - \left(\nabla_{\theta}^2 \ell(\theta) \right)^{-1} \nabla_{\theta} \ell(\theta)$$

1.2.1 Régression linéaire

On suppose ici que $y|x; \theta \sim \mathcal{N}(\mu, \sigma^2)$

□ **Équations normales** – En notant X la matrice de design, la valeur de θ qui minimize la fonction de cost a une solution de forme fermée tel que :

$$\theta = (X^T X)^{-1} X^T y$$

□ **Algorithme LMS** – En notant α le taux d'apprentissage, la règle de mise à jour d'algorithme des moindres carrés (LMS) pour un jeu de données d'entraînement de m points, aussi connu sous le nom de règle de Widrow-Hoff, est donné par :

$$\forall j, \quad \theta_j \leftarrow \theta_j + \alpha \sum_{i=1}^m \left[y^{(i)} - h_{\theta}(x^{(i)}) \right] x_j^{(i)}$$

Remarque : la règle de mise à jour est un cas particulier de l'algorithme du gradient.

□ **LWR** – Locally Weighted Regression, souvent noté LWR, est une variante de la régression linéaire appliquant un coefficient à chaque exemple dans sa fonction de coût via $w^{(i)}(x)$, qui est défini avec un paramètre $\tau \in \mathbb{R}$ de la manière suivante :

$$w^{(i)}(x) = \exp \left(- \frac{(x^{(i)} - x)^2}{2\tau^2} \right)$$

1.2.2 Classification et régression logistique

□ **Sigmoïde** – La sigmoïde g , aussi connue sous le nom de fonction logistique, est définie par :

$$\forall z \in \mathbb{R}, \quad g(z) = \frac{1}{1 + e^{-z}} \in]0,1[$$

□ **Régression logistique** – On suppose ici que $y|x; \theta \sim \text{Bernoulli}(\phi)$. On a la forme suivante :

$$\phi = p(y = 1|x; \theta) = \frac{1}{1 + \exp(-\theta^T x)} = g(\theta^T x)$$

Remarque : il n'y a pas de solution fermée dans le cas de la régression logistique.

□ **Régression softmax** – Une régression softmax, aussi appelée un régression logistique multi-classe, est utilisée pour généraliser la régression logistique lorsqu'il y a plus de 2 classes à prédire. Par convention, on fixe $\theta_K = 0$, ce qui oblige le paramètre de Bernoulli ϕ_i de chaque classe i à être égal à :

$$\phi_i = \frac{\exp(\theta_i^T x)}{\sum_{j=1}^K \exp(\theta_j^T x)}$$