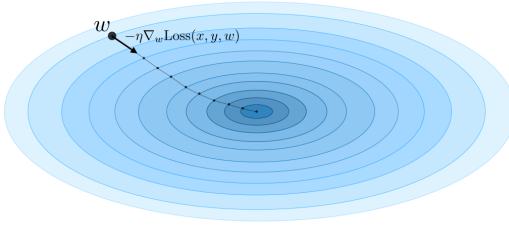


$\text{Loss}(x, y, w)$ de la manière suivante :

$$w \leftarrow w - \eta \nabla_w \text{Loss}(x, y, w)$$



□ **Mises à jour stochastiques** – L’algorithme du gradient stochastique (en anglais *stochastic gradient descent* ou *SGD*) met à jour les paramètres du modèle en parcourant les exemples $(\phi(x), y) \in \mathcal{D}_{\text{train}}$ de l’ensemble d’entraînement un à un. Cette méthode engendre des mises à jour rapides à calculer mais qui manquent parfois de robustesse.

□ **Mises à jour par lot** – L’algorithme du gradient par lot (en anglais *batch gradient descent* ou *BGD*) met à jour les paramètres du modèle en utilisant des lots entiers d’exemples (e.g. la totalité de l’ensemble d’entraînement) à la fois. Cette méthode calcule des directions de mise à jour des coefficients plus stable au prix d’un plus grand nombre de calculs.

1.5 Peaufinage de modèle

□ **Classe d’hypothèses** – Une classe d’hypothèses \mathcal{F} est l’ensemble des prédicteurs candidats ayant un $\phi(x)$ fixé et dont le paramètre w peut varier :

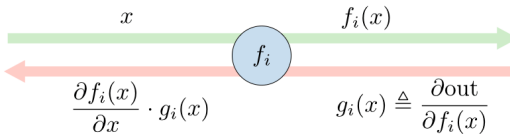
$$\mathcal{F} = \{f_w : w \in \mathbb{R}^d\}$$

□ **Fonction logistique** – La fonction logistique σ , aussi appelée en anglais *sigmoid function*, est définie par :

$$\forall z \in]-\infty, +\infty[, \quad \sigma(z) = \frac{1}{1 + e^{-z}}$$

Remarque : la dérivée de cette fonction s’écrit $\sigma'(z) = \sigma(z)(1 - \sigma(z))$.

□ **Rétropropagation du gradient (en anglais *backpropagation*)** – La propagation avant (en anglais *forward pass*) est effectuée via f_i , valeur correspondant à l’expression appliquée à l’étape i . La propagation de l’erreur vers l’arrière (en anglais *backward pass*) se fait via $g_i = \frac{\partial \text{out}}{\partial f_i}$ et décrit la manière dont f_i agit sur la sortie du réseau.



□ **Erreur d’approximation et d’estimation** – L’erreur d’approximation ϵ_{approx} représente la distance entre la classe d’hypothèses \mathcal{F} et le prédicteur optimal g^* . De son côté, l’erreur