



3I026 - INTRODUCTION À L'INTELLIGENCE ARTIFICIELLE ET AUX DATA SCIENCES

Vincent Guigue
Christophe Marsala

Sorbonne Université



- 1 Acquisition des données
- 2 Pré-traitements
- 3 Algorithmes d'apprentissage
- 4 Evaluation des performances

⇒ 1 étape = une fonction / un objet...

... Et des paramètres !

ATTENTION aux limites des notebooks

1 Acquisition des données

- Plus ou moins de données
- Switcher entre les bases (jouets / réelles)

2 Pré-traitements

- Activer / désactiver des fonctions de transformations
- Sauvegarder les résultats intermédiaires (`pickle`)

3 Algorithmes d'apprentissage

- Choix des algos (PObj = passer d'un classifieur à l'autre)

4 Evaluation des performances

- Boucle `grid-search` : détermination des paramètres optimaux

⇒ 1 étape = une fonction / un objet...

... Et des paramètres !

ATTENTION aux limites des notebooks

Fonction linéaire

$$\mathbf{x} = [x_1, \dots, x_d]$$

$$f(\mathbf{x}) = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_d x_d \text{ (+}\alpha_0\text{)}$$

- Je formule une fonction de coût...

un écart entre les étiquettes à prédire et la sortie de notre prédicteur :

$$\mathcal{L} = \Delta(f(\mathbf{x}), y), \text{ Pour l'ensemble des données : } \sum_i \Delta(f(\mathbf{x}_i), y_i)$$

$$\mathbf{x} = [x_1, \dots, x_d]$$

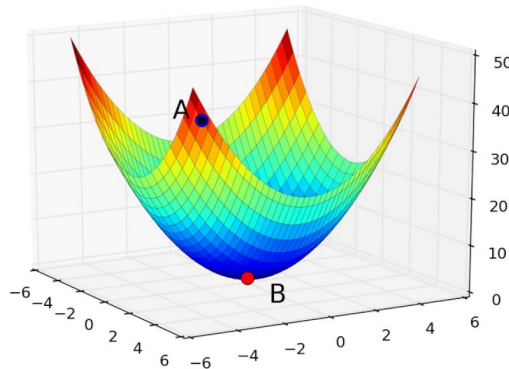
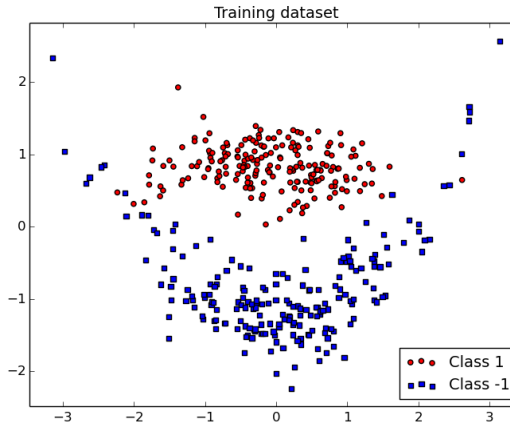
$$f(\mathbf{x}) = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_d x_d \text{ (+}\alpha_0\text{)}$$

- Je formule une fonction de coût...
un écart entre les étiquettes à prédire et la sortie de notre prédicteur :

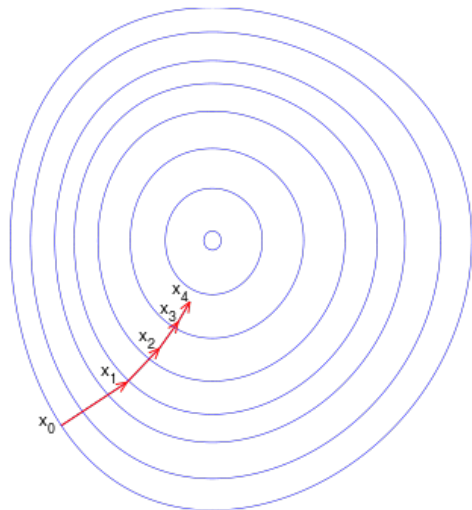
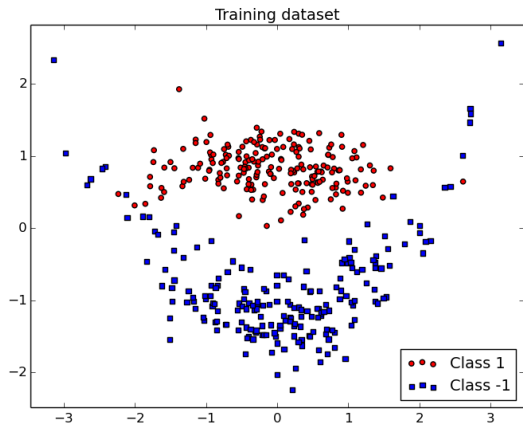
$$\mathcal{L} = \Delta(f(\mathbf{x}), y), \text{ Pour l'ensemble des données : } \sum_i \Delta(f(\mathbf{x}_i), y_i)$$

- Dérivation (=calcul du **gradient**)
 - Réflexion sur la dimension du gradient
- Annulation du gradient
 - ... Ou minimisation du gradient
- Signification de la dérivée d'une fonction
- Min vs ArgMin

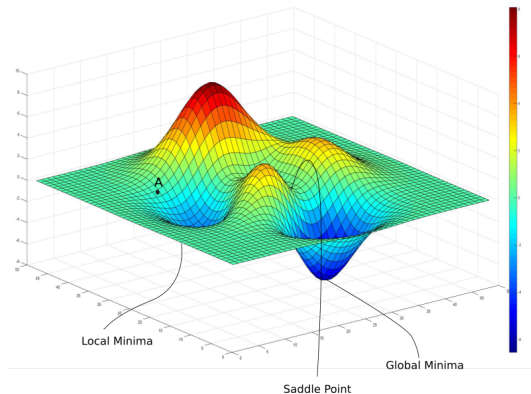
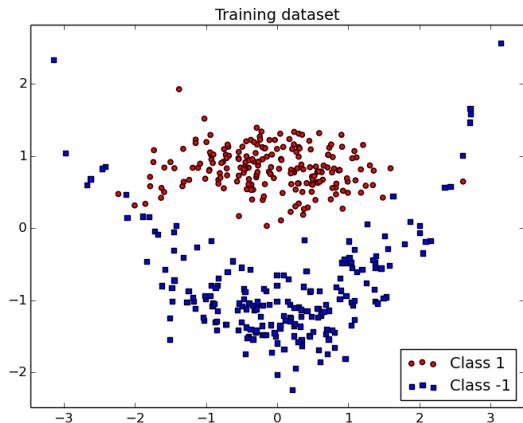
- Ne pas confondre les points et les paramètres



- Ne pas confondre les points et les paramètres



- Ne pas confondre les points et les paramètres



Différentes fonctions de coût

- Définition
- Calcul du gradient
- Résolution analytique
- Algorithme de la descente de gradient
- Algorithme de la descente de gradient stochastique

- Des coûts de classification mal maîtrisés

- Formulation
- Calcul du gradient
- Algorithme du perceptron
- Ajout d'une marge... Et d'une régularisation

Régularisation

Principe du rasoir d'Ockham

Application sur le perceptron

- En théorie (résultat attendu)
- En pratique (implémentation)

- Normalisation du gradient
- Réduction du pas d'apprentissage ϵ
- Ajout d'une marge sur le perceptron