

ARF

Regression linéaire Descente de gradient

Cours 3

Nicolas Baskiotis

`nicolas.baskiotis@lip6.fr`

Master 1 DAC

équipe MLIA, Laboratoire d'Informatique de Paris 6 (LIP6)
Université Pierre et Marie Curie (UPMC)

S2 (2014-2015)

Plan

1 **Regression linéaire**

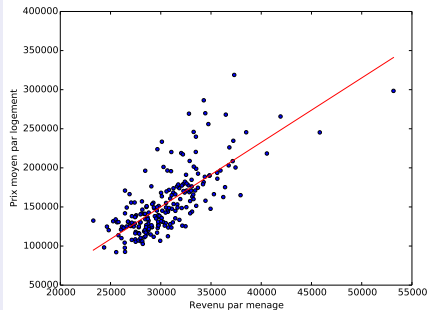
2 Régression logistique

3 Descente de gradientt

Introduction

Regression linéaire

- Objectif : prédire une sortie continue réelle y à partir d'un nombre de variables d'entrée
- beaucoup d'applications, très utilisée un peu dans tous les domaines
- très flexible (transformation des entrées)



Formalisation

Objectifs

Etant donné un ensemble $\{(x^i, y^i)\} \in \mathbb{R}^d \times \mathbb{R}$,

- on cherche une sortie linéaire en fonction des entrées :

$$E(y|x) = w_0 + \sum_{i=1}^d w_i x_i$$

$$\Rightarrow f(x) = w_0 + \sum_{i=1}^d w_i x_i$$

- qui fait le moins d'erreurs : $f(x_i)$ doit être proche de y_i
- sous la condition que l'erreur est indépendante de x , de variance σ^2 constante, suit une loi normale.

$\Rightarrow y|x \sim \mathcal{N}(f(x), \sigma^2)$ (lien avec l'apprentissage bayésien)

- Erreur moindres carrés (MSE) : $\sum_{i=1}^n (y^i - f(x^i))^2$

Résolution

Méthode

- Minimiser $\sum_{j=1}^n (y^j - f(x^j))^2 = \sum_{j=1}^n (y^j - w_0 - \sum_{i=1}^d w_i x_i^j)^2$
- Annuler le gradient \rightarrow solution analytique.

Ecriture matricielle

- $$X = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix} = \begin{pmatrix} x_1^1 & x_2^1 & \dots & x_d^1 \\ x_1^2 & x_2^2 & \dots & x_d^2 \\ & & \vdots & \\ x_1^n & x_2^n & \dots & x_d^n \end{pmatrix}, Y = \begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^n \end{pmatrix}, W = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix}$$

- Minimiser : $(Y - W'X)(Y - W'X)'$
- Solution : $(X'X)^{-1}X'Y$
- Et pour w_0 ?

Plan

1 Régression linéaire

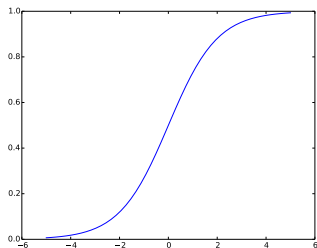
2 Régression logistique

3 Descente de gradientt

Problématique

Comment adapter la régression à la classification ?

- Régression : $y|x \sim \mathcal{N}(w^t x, \sigma^2)$
 - Pas adapter ! On veut une classification binaire
- ⇒ Modélisation par une variable de Bernoulli :
- $$p(y = 1|x) = \mu(x)^y (1 - \mu(x))^{1-y}$$
- Comment représenter $\mu(x)$? fonction linéaire ?
 - Non ... (pourquoi ?)
- ⇒ fonction logistique : $\mu(x) = \frac{1}{1+e^{-\eta(x)}}$



Régression logistique

Principe

Quand est-ce que :

- $p(+|x) = 0.5$?
- $p(+|x) < 0.5$?
- $p(+|x) > 0.5$?

Résolution

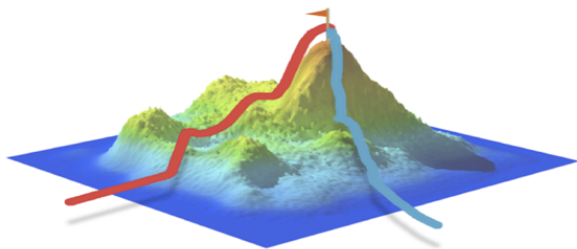
- Pas de solution analytique.
- Méthode d'optimisation numérique \rightarrow descente de gradient.

Plan

- 1 Régression linéaire
- 2 Régression logistique
- 3 Descente de gradientt**

Principe

- Algorithme d'optimisation différentiable
- S'applique pour toute fonction différentiable
- Idée simple : améliorer de façon itérative la solution courante



Quelques notations et rappels

Convexité

- C ensemble convexe de \mathbb{R}^n : $\forall x, y \in C, \forall \lambda \in [0, 1], \lambda x + (1 - \lambda)y \in C$
 - $\sum_i \lambda_i x_i$ est une combinaison convexe ssi $\forall i, \lambda_i \geq 0$ et $\sum_i \lambda_i = 1$
 - Enveloppe convexe d'un ensemble fini $\{x_i\}, i = 1 \dots n$: toutes les combinaisons convexes de l'ensemble
 - Fonction convexe $f : X \rightarrow \mathbb{R}$ ssi $\forall x, x' \in X, \forall \lambda \in [0, 1]$ tq $\lambda x + (1 - \lambda)x' \in X$ alors $f(\lambda x + (1 - \lambda)x') \leq \lambda f(x) + (1 - \lambda)f(x')$
- si $\lambda_i \geq 0$ et $\sum_i \lambda_i = 1$, alors $f(\sum_i \lambda_i x_i) \leq \sum_i \lambda_i f(x_i)$ (inégalité de Jensen)

Quelques notations et rappels

Différentiabilité

- Si $f : X \rightarrow \mathbb{R}$ est convexe ssi $\forall x, x' \in X, f(x') \geq f(x) + \langle x' - x, \nabla f(x) \rangle$
- Si f convexe, alors sa matrice hessienne est définie semi-positive : $\nabla^2 f \geq 0$.

Minimum

- Si f atteint son minimum, alors les minimums forment un ensemble convexe.
- Si l'ensemble est strictement convexe, le minimum est un singleton.

Algorithme du gradient

Algorithme

- 1 Choisir un point x_0
- 2 Itérer :
 - ▶ Calculer $\nabla f(x_t)$
 - ▶ mettre à jour $x_{t+1} \leftarrow x_t - \alpha \nabla f(x_t)$

