1. Régression linéaire et non linéaire régularisée

1.1 Régression linéaire:

1. L'ensemble θ des paramètres est {w, b} : w étant le vecteur de poids de dimension R^d et b, le biais de dimension R.

2. Le risque empirique est $\hat{R}(f_\theta, D_n) = \sum_{i=1}^{n} L\left(f_\theta\left(x^{(i)}\right), t^{(i)}\right)$

3. Pour minimiser le risque empirique, on cherche le θ qui donne le moins d'erreur sur l'ensemble d'entraînement, soit : $\theta^* = arg\ \min_\theta \hat{R}(f_\theta, D_n)$

4. Le gradient du risque empirique est :
$\nabla\hat{R}(f_\theta, D_n) = \sum_{i=1}^{n} \frac{d}{d\theta} L\left(f_\theta\left(x^{(i)}\right), t^{(i)}\right) = \sum_{i=1}^{n} \frac{d}{d\theta} f\left(\left(x^{(i)}\right) - t^{(i)}\right)^2$

1.2 Régression linéaire régularisée ("ridge regression"):

1. Le gradient du risque régularisé est :
$\nabla\tilde{R}_\lambda(f_\theta, D_n) = \sum_{i=1}^{n} \frac{d}{d\theta} L\left(f_\theta\left(x^{(i)}\right), t^{(i)}\right) + \lambda \frac{d}{d\theta}\Omega(\theta) = \sum_{i=1}^{n} \frac{d}{d\theta} f\left(\left(x^{(i)}\right) - t^{(i)}\right)^2 + \lambda \frac{d}{d\theta}\Omega(\theta)$

   Expliquer la différence avec le risque non régularisé

2. DescenteDeGradientBatch(ϵ, η, θ_0)
   $\theta \leftarrow \theta_0$
   faire
      $\theta \leftarrow \theta - \eta\nabla\tilde{R}_\lambda$
   jusqu'à | $\eta\nabla\tilde{R}_\lambda$ | < ϵ

1.3 Régression avec un pré-traitement non-linéaire fixe:

1. $\tilde{f}(x) = f\left(\phi_{poly^k}(x)\right) = f\begin{pmatrix} x \\ x^2 \\ \vdots \\ x^k \end{pmatrix}$

2. k ϵ $\mathbb{N}$ est de dimension 1

3. Avec x en dimension d = 2, on a:
$\phi_{poly^1}(x) = (x_1\ \ x_2)^T$
$\phi_{poly^2}(x) = (x_1\ \ x_2\ \ {x_1}^2\ \ x_1x_2\ \ {x_2}^2)^T$
$\phi_{poly^3}(x) = (x_1\ \ x_2\ \ {x_1}^2\ \ x_1x_2\ \ {x_2}^2\ \ {x_1}^3\ \ {x_1}^2{x_2}^1\ \ {x_1}^1{x_2}^2\ \ {x_2}^3)^T$

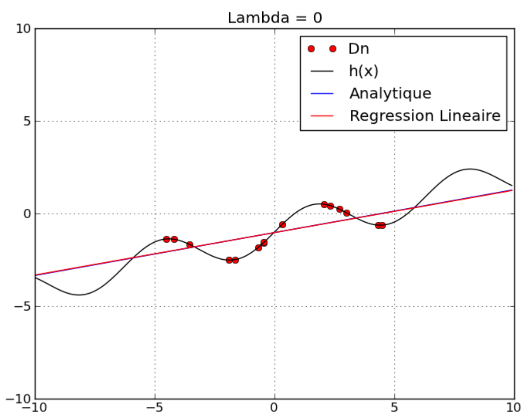
$$\phi_{poly^4}(x) = (x_1\ x_2\ x_1{}^2\ x_1x_2\ x_2{}^2\ x_1{}^3\ x_1{}^2x_2{}^1\ x_1{}^1x_2{}^2\ x_2{}^3\ x_1{}^4\ x_1{}^3x_2{}^1\ x_1{}^2x_2{}^2\ x_1{}^1x_2{}^3\ x_2{}^4)^T$$

4. Avec x en dimension d, on à $\phi_{poly^k}(x)$ de dimension $\sum_{n=1}^{k} \frac{(n+d-1)!}{n!\,(d-1)!}$

Partie 2: Programmation

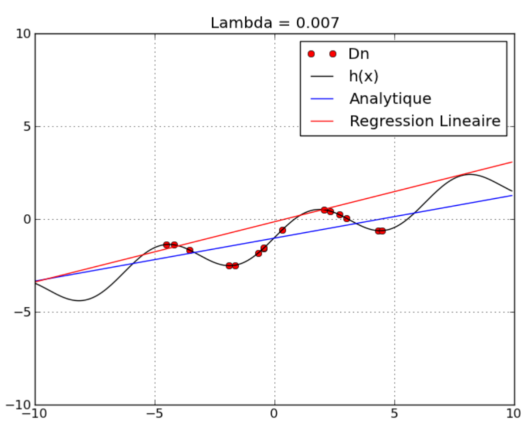
4)

Lambda = 0
Paramètres :
Analytique : w = 0.23179125, b = -1.02987640
Régression: w = 0.22932296, b = -1.03025983

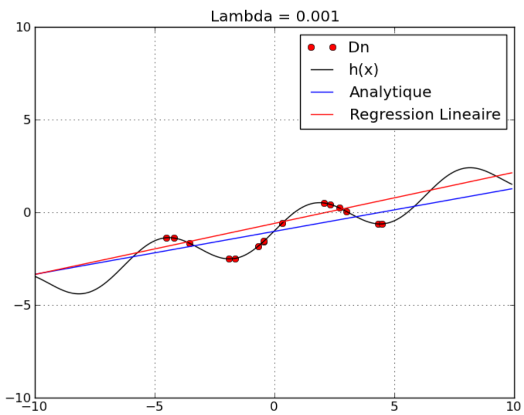


5)

Valeurs numériques pour un lambda (valeur extrême) = 0.007
Analytique : w = 0.23177799, b = -1.02987477
Régression: w = 0.32455948, b = -0.14699578

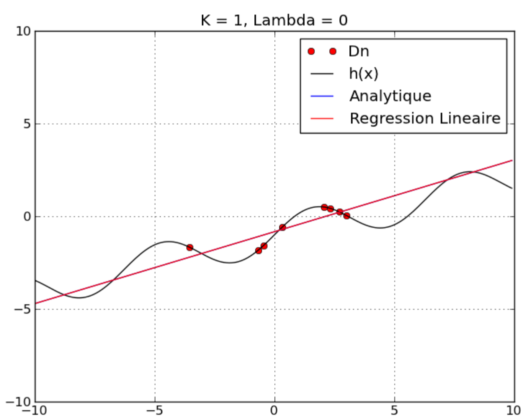


Valeurs numériques pour un lambda (valeur intermédiaire) = 0.001
Analytique : w = 0.23178936, b = -1.02987617
Régression: w = 0.27562477, b = -0.60097945

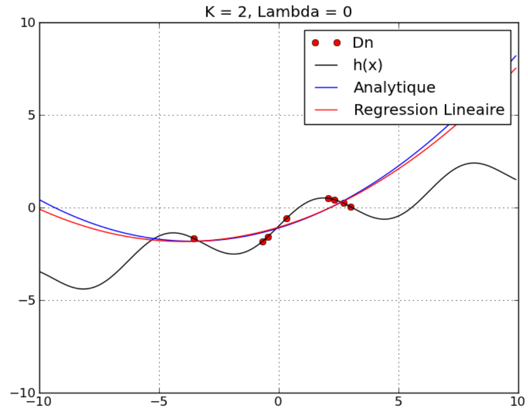


6)

k = 1
Regression Analytique : [-0.85256288 0.38943102]
Regression Gradient : [0.38945179 -0.8505361]



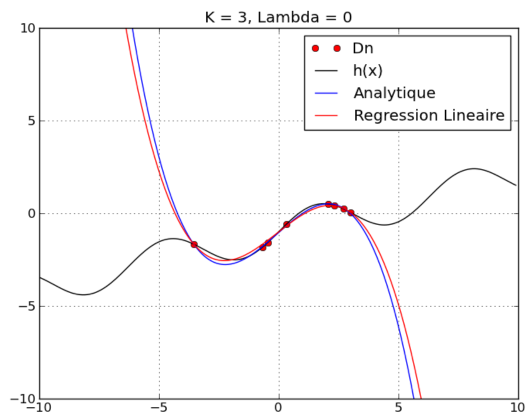
k = 2
Regression Analytique : [-1.27421837 0.4044939 0.07332763]
Regression Gradient : [0.38960116 0.0574411 -1.14491271]



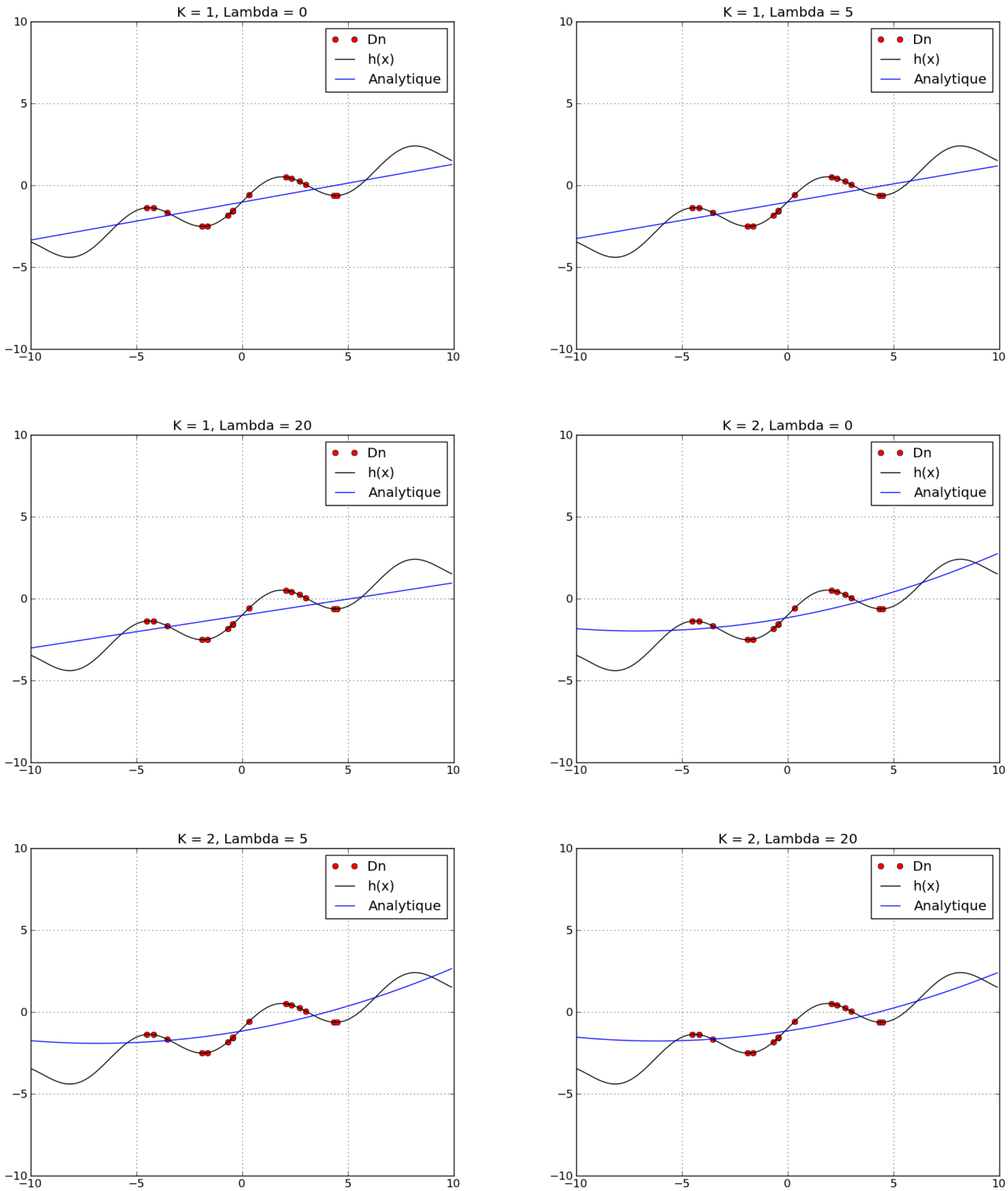
k = 3

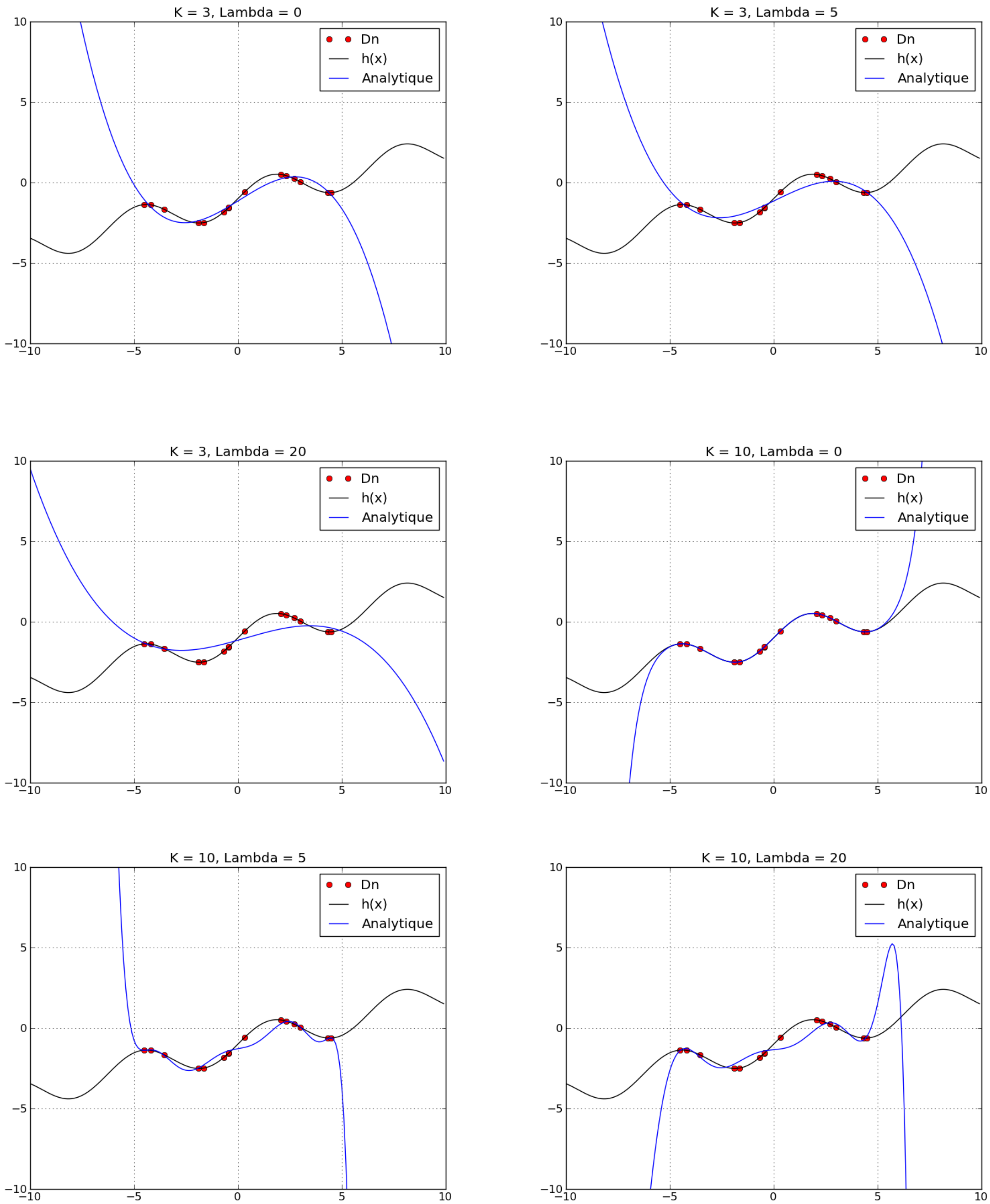
Regression Analytique : [-1.08185509 1.11511891 -0.01274914 -0.07913959]

Regression Gradient : [1.07271974 -0.01828174 -0.07517304 -0.9904767]

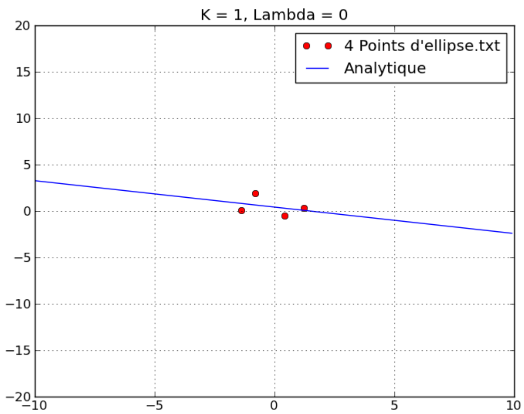


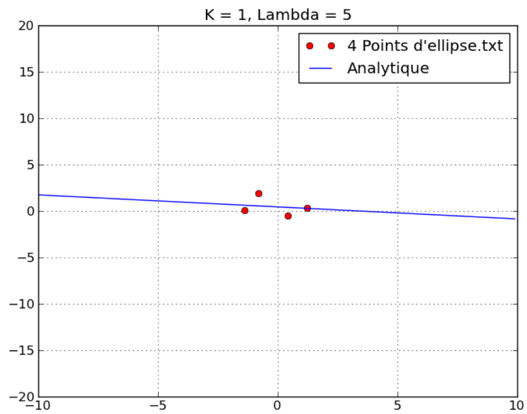
7) Effets de la variation de lambda pour un K donné (K = 1,2,3,10 ; Lambda = 0,5,20)
-> Expliquer l’influence de l’hyper-paramètre lambda

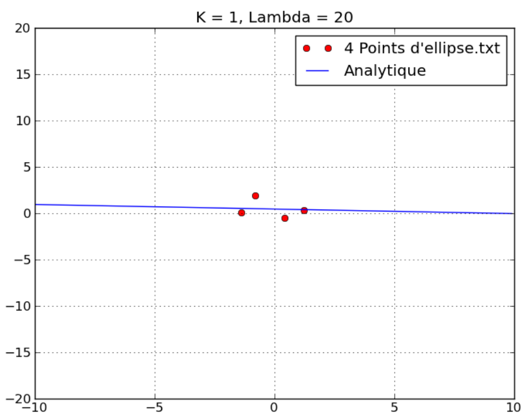


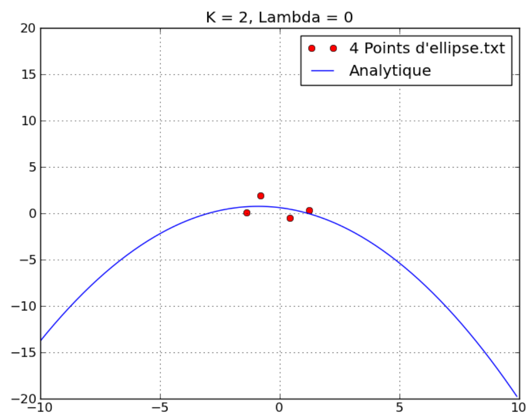

K = 3, Lambda = 0
Dn
h(x)
Analytique
K = 3, Lambda = 5
Dn
h(x)
Analytique
K = 3, Lambda = 20
Dn
h(x)
Analytique
K = 10, Lambda = 0
Dn
h(x)
Analytique
K = 10, Lambda = 5
Dn
h(x)
Analytique
K = 10, Lambda = 20
Dn
h(x)
Analytique

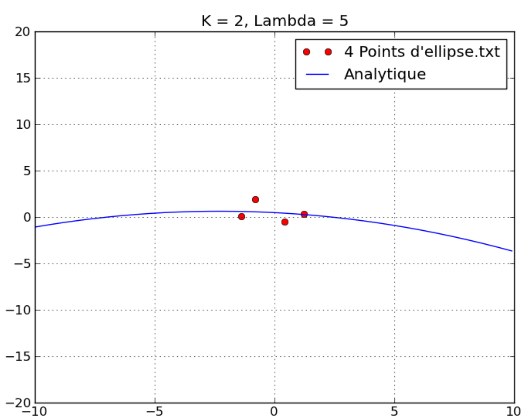
8) Nous appliquons maintenant l'algorithme de regression_analytique sur le problème de classification 2D « ellipse.txt », avec des prétraitements k = 1, 2, 3, 4 pour des lambdas = 0, 5 , 20

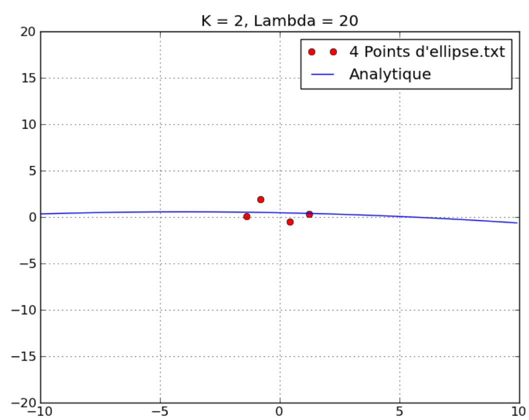


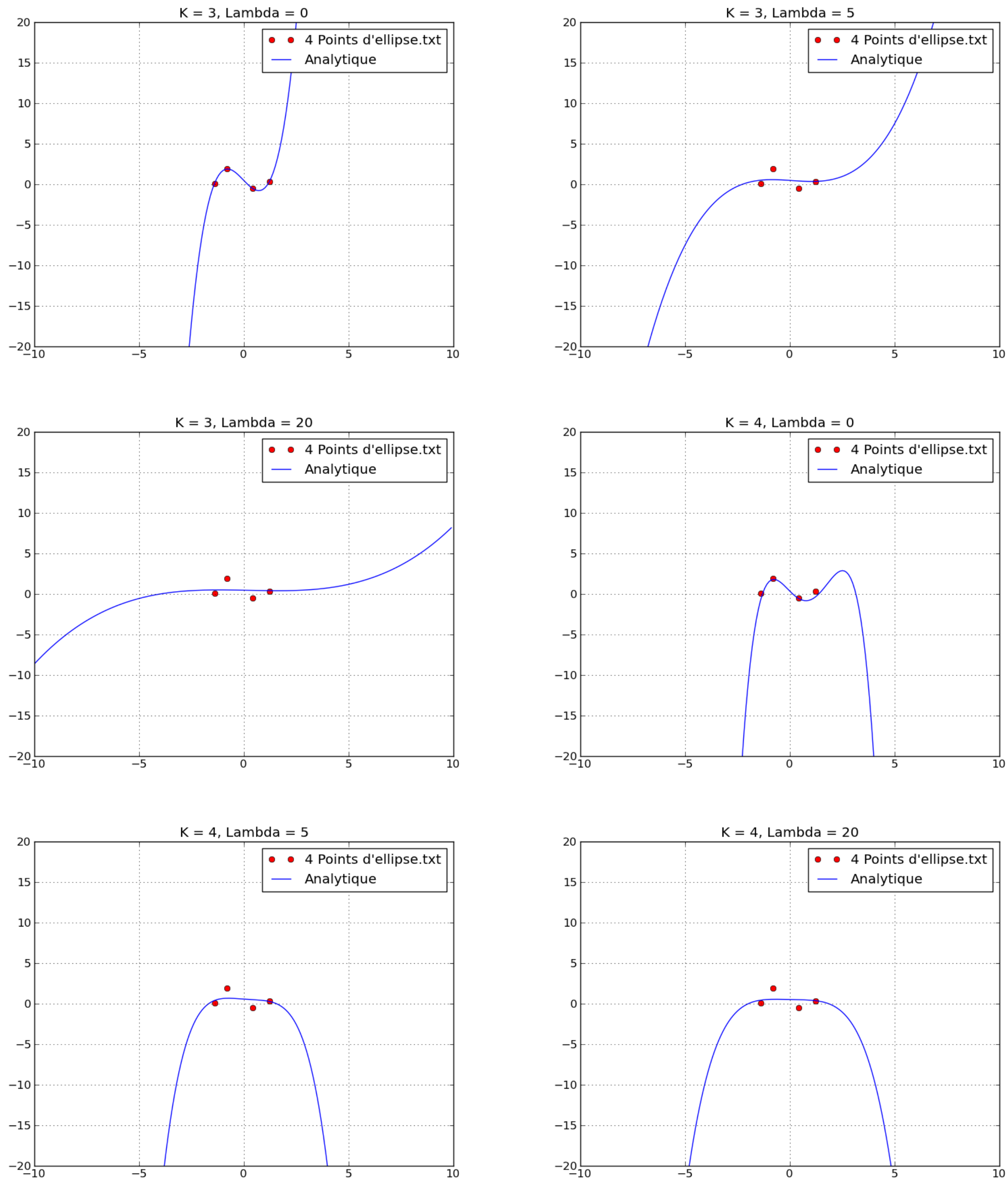











K = 3, Lambda = 0
4 Points d'ellipse.txt
Analytique
K = 3, Lambda = 5
4 Points d'ellipse.txt
Analytique
K = 3, Lambda = 20
4 Points d'ellipse.txt
Analytique
K = 4, Lambda = 0
4 Points d'ellipse.txt
Analytique
K = 4, Lambda = 5
4 Points d'ellipse.txt
Analytique
K = 4, Lambda = 20
4 Points d'ellipse.txt
Analytique