

Risque réel et risque empirique

Bruno Bouzy

6 février 2013

Introduction

Ce document illustre par l'exemple ce qu'est le « risque réel » au sens de [1] et, dans une moindre mesure, ce qu'est le « risque empirique ». Pour résumer [1], lorsque un système apprend, il le fait dans un espace d'hypothèses distinct de l'espace réel. L'espace réel n'est modélisé, formalisé que par l'espace des hypothèses. En apprenant de cette manière, le système commet deux erreurs: l'erreur d'approximation ou « biais »: utiliser un espace d'hypothèses plutôt que l'espace réel, et l'erreur d'estimation ou « variance »: trouver une hypothèse qui n'est malheureusement pas optimale. Le « risque réel » est la somme des erreurs faites, biais plus variance. Le risque réel, et plus particulièrement le biais, sont difficilement mesurables puisque la modélisation est faite dans l'espace des hypothèses. Dans l'espace des hypothèses, on mesure l'erreur d'estimation ou « risque empirique ». Une approche de l'apprentissage automatique est de minimiser le risque empirique. Le risque empirique se mesure sur un échantillon S de taille N . Si S est l'échantillon sur lequel le système apprend, et si N est petit alors le risque empirique peut être nul. Lorsque N augmente le risque empirique augmente car le système a plus de difficultés à apprendre. En revanche, lorsque N augmente, le risque réel est censé diminuer.

Le problème

Pour illustrer l'apprentissage automatique, nous allons choisir le problème de l'approximation de fonctions f définies sur $[0,1]$ et nous allons prendre pour espace des hypothèses l'ensemble des fonctions h constantes par intervalles.

La figure 1 représente la fonction f définie par $f(x) = 1-x$ sur l'intervalle $[0,1]$.

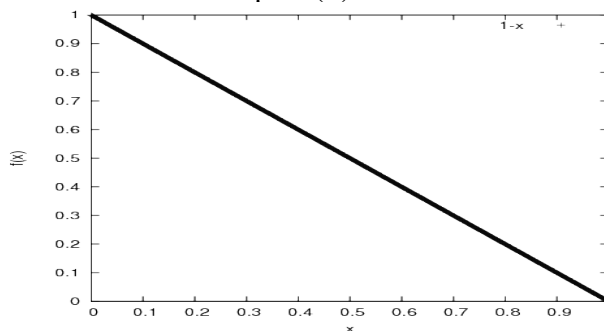


Figure 1: la fonction $f(x) = 1-x$ sur l'intervalle $[0,1]$.

Dans le cadre de l'apprentissage automatique, il faut considérer que l'on ne connaît pas la définition de cette fonction f mais que l'on peut demander la valeur $f(x)$ de la fonction sur un échantillon S de points.

Soit N la taille de l'échantillon. On écrit $[0, 1[= \bigcup_{0 \leq i < N} [i/N, (i+1)/N[$. On demande la valeur de f pour les N valeurs de x de la forme $(i+0.5)/N$ avec $0 \leq i < N$. Pour les autres points x de $[0, 1[$, on définit $h_N(x)$ de la manière suivante: soit i l'entier tel que $i \leq xN < (i+1)$, on définit $h_N(x)$

= $f((i+0.5)/N)$. h_N est donc une fonction constante sur les intervalles de la forme $[i/N, (i+1)/N[$. Avec cette définition, on constate que l'erreur d'estimation est nulle sur l'ensemble d'apprentissage: $f(x)=h_N(x)$ pour $x=(i+0.5)/N$. Le risque réel ou erreur totale est donné théoriquement par (1).

$$\text{risque réel} = \int_0^1 |f(t)-h_N(t)| dt \quad (1)$$

Pour le calculer, on va prendre $M>N$ et l'approximer avec (2).

$$\text{risque réel} = \left(\sum_{0 \leq i < M} |f((i+0.5)/M) - h_N((i+0.5)/M)| \right) / M \quad (2)$$

Pour être capable de mesurer et visualiser le risque réel en fonction de N , nous allons supposer que l'espace réel n'est pas celui des fonctions sur $]0,1[$ mais des fonctions sur une discrétisation de $]0,1[$ avec une précision suffisante.

Les fonctions à approximer

Pour illustrer le risque réel, on va prendre plusieurs fonctions:

- une fonction « facile » $f(x)=1-x$,
- une fonction « moyenne » f_{permute} ,
- une fonction « difficile » f_{decale} ,
- une fonction « impossible » f_{randon} .

La fonction f_{permute} prend x en écriture décimale et permute les deux premières décimales de x . Exemples:

$$\begin{array}{lll} f(0) = 0 & f(0,1234) = 0,2134 & f(0,5) = 0,05 \\ f(0,99) = 0,99 & f(0,676767) = 0,766767 & f(0,9111) = 0,1911 \end{array}$$

La figure 2 montre la fonction f_{permute} . f_{permute} est continue et affine par intervalle, mais n'est pas continue sur $[0,1[$.

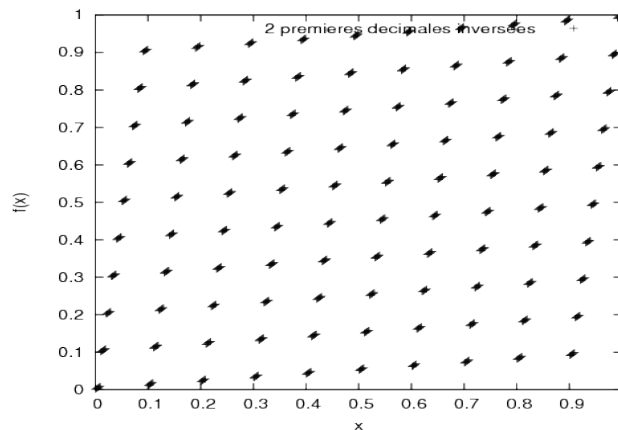


Figure 2: f_{permute} .

La fonction f_{decale} prend x en écriture décimale. Pour connaître la valeur de la i ème décimale de $f(x)$, f lit la i ème décimale de x di et définit $j=i+di/3$ et définit que la i ème décimale de $f(x)$ est la j ème décimale de x . Exemples:

$$\begin{array}{lll} f(0,123456789) = 0,1245689 & f(0) = 0 & f(0,5) = 0 \\ f(0,99) = 0 & f(0,676767) = 0,6767 & f(0,9111) = 0,1111 \end{array}$$

La figure 3 montre la fonction f_{decale} . f_{decale} est compliquée!

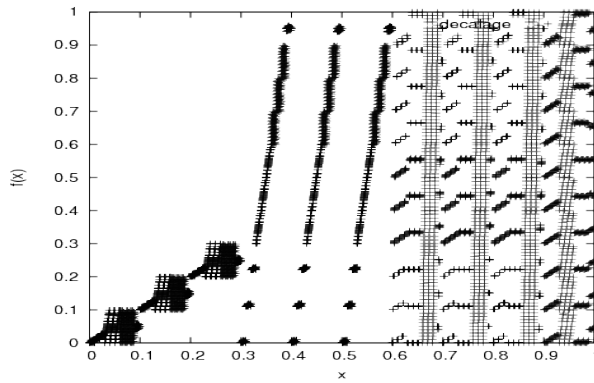


Figure 3: *fdecale*.

Enfin, la fonction *frandom* est la fonction aléatoire. Pour chaque valeur de x , une valeur de $f(x)$ est tirée au hasard dans $[0,1[$. La figure 4 montre la fonction *frandom*.

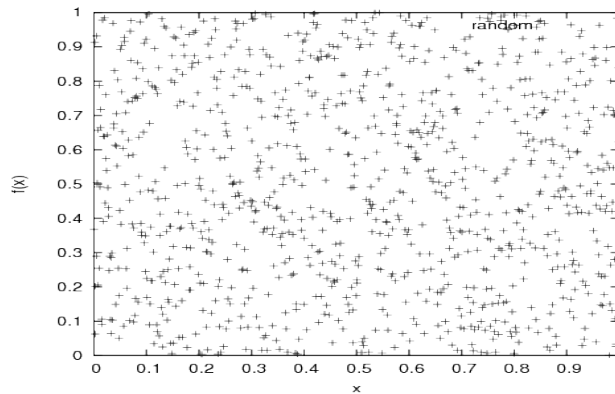


Figure 4: *frandom*.

Mesure du risque réel

Pour chacune des fonctions, on a mesuré le risque réel avec $M = 100000$. La figure 5 montre les courbes de risque réel pour les 4 fonctions lorsque N est une puissance de 2 inférieure à M .

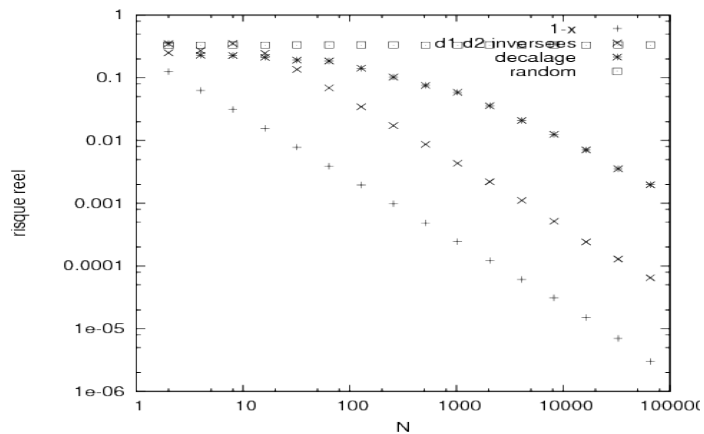


Figure 4: Le risque réel.

Premièrement, pour N donné, on observe que plus la fonction à approximer est « difficile », plus le risque réel mesuré est élevé.

On observe que le risque réel de la fonction $1-x$ décroît rapidement: avec un échantillon de taille 10^5 on observe un risque réel inférieur à 10^{-5} . Le risque réel varie en $1/N$.

Pour la fonction *fpermute*, les discontinuités de la fonction rendent le risque réel constant pour les valeurs de N inférieures à 100 environ. Pour N supérieur à 100, le risque réel décroît. La fonction étant continue par morceaux, le risque réel tend vers 0.

Pour la fonction *fdecale*, on observe que le risque réel diminue lorsque N augmente, mais plus lentement que pour les deux fonctions précédentes. La i ème décimale de $f(x)$ ne dépend que de la i ème décimale de x et des 3 décimales suivantes. Si x n'est pas proche d'un point de discontinuité de *fdecale*, et si on veut connaître les i premières décimales de $f(x)$, il suffit de connaître les $i+3$ premières décimales de x . Autrement dit, si on veut connaître $f(x)$ avec une précision 10^{-p} , il suffit de connaître x avec une précision 10^{-p-3} . Avec N suffisamment grand on peut rendre le risque réel aussi petit que l'on veut, et le risque réel tend vers 0.

Pour les 3 fonctions précédentes, on constate que le biais est nul (le risque réel tend vers 0). L'espace des hypothèses avec des fonctions constantes par intervalle est une bonne modélisation.

Enfin, pour la fonction *frandom*, on observe que le risque réel est constant. Quel que soit la petitesse des intervalles où h est constante, on trouve des valeurs de f suffisamment loin de celle de h , et h n'arrive pas à approximer f . Pour *frandom*, le biais observé est strictement positif et constant (≈ 0.3).

Conclusion

On a illustré le risque réel pour 4 fonctions à approximer. Comme dans les bons cas d'apprentissage de [1], on a observé que le risque réel décroît lorsque N augmente (sauf pour *frandom*). On a observé un biais nul pour 3 fonctions et un biais strictement positif pour la dernière. On a observé un risque empirique nul pour les 4 fonctions.

Références

[1] Antoine Cornuéjols, Laurent Miclet, Apprentissage artificiel, concepts et algorithmes, Eyrolles, 2003.