

Traitement du signal et ses applications

Reconnaissance des formes et machine learning

R. Flamary, A. Rakotomamonjy, C. Richard

8 avril 2013



Qu'est ce que la reconnaissance de forme ?

Quelques définitions provenant de la littérature

- ▶ le processus d'affectation d'un **objet physique ou un évènement** à une ou plusieurs **catégories** pré-spécifiées (*Duda et Hart*).
- ▶ Étant donné, plusieurs exemples de **signaux complexes** et d'étiquettes (ou décisions) associées, la RdF est le processus de prise de décision automatique pour un ensemble d'autres signaux. *Ripley*
- ▶ Le processus d'affectation d'un **nom** w à une **observation** x . *Schurmann*

But de la reconnaissance de forme

Permettre à la machine de traiter automatiquement des masses de données (signaux, images) pour résoudre un problème donné.



Plan du cours

Introduction

- Exemples de problèmes
- Types de problèmes RdF
- Définitions

Description/Exploration des données

- Clustering
- Estimation de densité de probabilité
- Réduction de dimension/ Visualisation

Prédiction

- Discrimination/Classification
- Régression

Mise en oeuvre d'un système de RdF

- Données réelle
- Sélection de modèles et de paramètres
- Exemples de mise en oeuvre



Exemples de problèmes de reconnaissance de formes

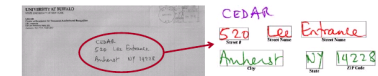
Vision

- ▶ Inspection de pièce
- ▶ Cibles militaires



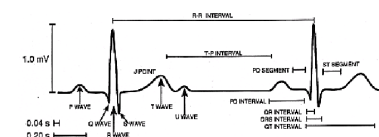
Reconnaissance de caractères

- ▶ Classement de courrier
- ▶ Traitement de chèques



Aide au diagnostic

- ▶ Imagerie médical, EEG, ECG
- ▶ Pour assister les médecins (et non les remplacer)



Types de problèmes RdF

Apprentissage non-supervisé

- ▶ **Clustering** Organiser les objets en des groupes présentant une certaine similarité (taxonomie des espèces animales).
- ▶ **Estimation de densité de probabilité** Estimer la loi de probabilité des données d'apprentissage (estimer la loi d'un bruit).
- ▶ **Réduction de dimension** Diminuer la dimensionnalité des données pour pouvoir mieux les interpréter/visualiser (recommandation).

Apprentissage supervisé

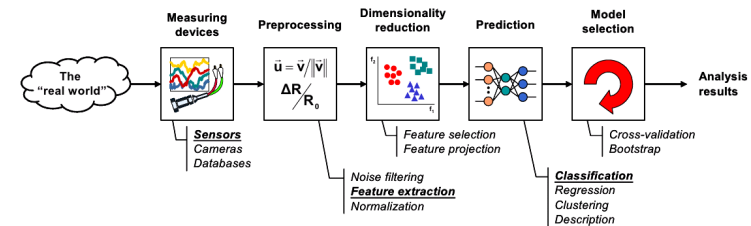
- ▶ **Discrimination/Classification** Affecter une classe à une observation (reco. de caractère, météo pluie).
- ▶ **Régression** Prédire une valeur réelle à partir d'une observation (météo température).



Les composantes d'un système de RdF

Un système classique est composé de

- ▶ d'un capteur
- ▶ d'un ensemble de pré-traitement des signaux
- ▶ d'un système d'extraction de caractéristiques
- ▶ d'un algorithme de classification
- ▶ d'un ensemble d'exemples (les données d'apprentissage)



Données d'apprentissage

Apprentissage non-supervisé

- ▶ $\mathbf{x} \in \mathbb{R}^d$ est une observation de d caractéristiques réelles.
- ▶ L'ensemble d'apprentissage définit par les observations $\{\mathbf{x}_i\}_{i=1}^n$ où n est le nombre de points (d'exemples) d'apprentissages.
- ▶ Les exemples sont souvent mis sous la forme d'une matrice $\mathbf{X} \in \mathbb{R}^{n \times d}$ définie par $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T$ contenant les exemples d'apprentissage en lignes et les caractéristiques en colonnes.
- ▶ d et n définissent la dimensionnalité du problème d'apprentissage.

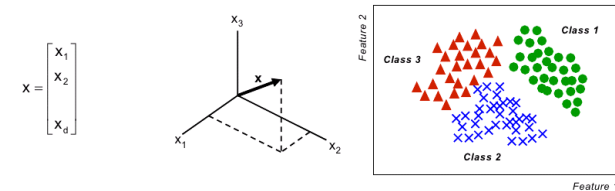
Apprentissage supervisé

- ▶ On associe à chaque observation \mathbf{x}_i une valeur à prédire $y_i \in \mathcal{Y}$.
- ▶ Tout comme pour les observation les valeurs à prédire (label) peuvent être concaténées en un vecteur $\mathbf{y} \in \mathcal{Y}^n$
- ▶ L'espace des valeurs à prédire \mathcal{Y} sera :
 - ▶ $\mathcal{Y} = \{-1, 1\}$ ou $\mathcal{Y} = \{1, \dots, m\}$ pour la classification.
 - ▶ $\mathcal{Y} = \mathbb{R}$ pour la régression.



Caractéristiques et formes

- ▶ Une **caractéristique** est un trait distinctif, ou caractéristique d'un objet. Il peut être **symbolique** (ex : une couleur) ou **numérique** (ex : taille).
- ▶ **Définition**
 - ▶ Une combinaison de caractéristiques est représentée à l'aide d'un vecteur \mathbf{x} de dimension d .
 - ▶ L'espace de dimension d contenant est appelé l'**espace de représentation**
 - ▶ Les objets sont représentés comme des points dans l'espace de représentation. On appelle cette représentation **diagramme de dispersion**



- ▶ Une **forme** est un ensemble de trait de caractéristiques d'une observation donnée. Dans les problèmes de discrimination, une forme est composée d'un **vecteur de caractéristiques** et d'un **label**



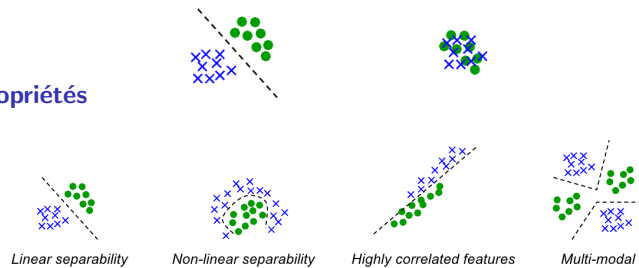
Caractéristiques

Qu'est ce qu'une "bonne" caractéristique ?

La qualité d'une caractéristique dépend du problème d'apprentissage.

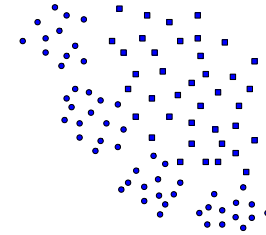
- ▶ **Discrimination** Les exemples d'une même classe devraient avoir des caractéristiques similaires alors que les exemples de classes différentes devraient avoir des caractéristiques différentes.
- ▶ **Régression** La caractéristique doit aider à mieux prédire la valeur (elle doit être corrélée avec les valeurs à prédire).

Autres propriétés



◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

Description/Exploration des données



Soit un ensemble d'apprentissage $\{\mathbf{x}_i\}_{i=1}^n$ composé d'exemples de dimension d

Objectifs

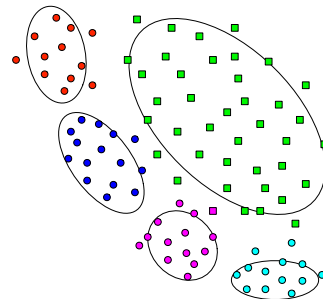
- ▶ **Clustering** $\{\mathbf{x}_i\}_{i=1}^n \Rightarrow \{\hat{y}_i\}_{i=1}^n$ où \hat{y} est une appartenance à un groupe.
- ▶ **Estimation de densité de probabilité** $\{\mathbf{x}_i\}_{i=1}^n \Rightarrow p(\mathbf{x})$.
- ▶ **Réduction de dimension** $\{\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n \Rightarrow \{\tilde{\mathbf{x}}_i \in \mathbb{R}^{d'}\}_{i=1}^n$ avec $d' \gg d$.

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

Clustering

Objectif

- ▶ Organiser les exemples d'apprentissage par groupes.
- ▶ $\{\mathbf{x}_i\}_{i=1}^n \Rightarrow \{\hat{y}_i\}_{i=1}^n$ où $\hat{y} \in \mathcal{Y}$ représente une classe ($\{1, \dots, m\}$)
- ▶ Paramètres :
 - ▶ m nombre de classes.
 - ▶ mesure de similarité



Exemples

- ▶ Taxonomie d'animaux.
- ▶ Regroupement de gènes.
- ▶ Réseaux sociaux.

Méthodes

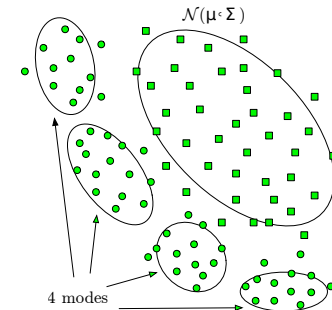
- ▶ k-means (k-moyennes).
- ▶ Mélange de gaussiennes.
- ▶ Clustering hiérarchique.

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

Estimation de densité de probabilité

Objectif

- ▶ Estimer la loi de proba des données.
- ▶ $\{\mathbf{x}_i\}_{i=1}^n \Rightarrow p(\mathbf{x})$ où $p(\mathbf{x})$ est une densité de proba ($\int p(\mathbf{x})d\mathbf{x} = 1$)
- ▶ Modèle peut être génératif.
- ▶ Paramètres :
 - ▶ Type de loi (gaussienne, ...).
 - ▶ Paramètres de la loi (μ, Σ)



Méthodes

- ▶ Fenêtres de Parzen.
- ▶ Histogramme.
- ▶ Mélange de gaussiennes.

Exemples

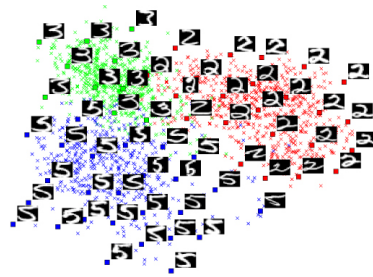
- ▶ Estimation de bruit.
- ▶ Génération de données (visage, ...).
- ▶ Détection de nouveauté.

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

Réduction de dimension

Objectif

- ▶ Projeter les données dans un espace de faible dimension.
- ▶ $\{\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n \Rightarrow \{\tilde{\mathbf{x}}_i \in \mathbb{R}^{d'}\}_{i=1}^n$ avec $d' \ll d$ (souvent $d' = 2$).
- ▶ Utilisation en visualisation, interprétation, pré-traitement des données.
- ▶ Paramètres :
 - ▶ Type de projection.
 - ▶ Mesure de similarité.



Méthodes

- ▶ Sélection de caractéristiques.
- ▶ Analyse en composantes principales (ACP, PCA).
- ▶ Réduction non-linéaire.

Exemples

- ▶ Visualisation de vecteurs.
- ▶ Interprétation des données.
- ▶ Systèmes de recommandation.



Section

Introduction

- Exemples de problèmes
- Types de problèmes RdF
- Définitions

Description/Exploration des données

- Clustering
- Estimation de densité de probabilité
- Réduction de dimension/ Visualisation

Prédiction

- Discrimination/Classification
- Régression

Mise en oeuvre d'un système de RdF

- Données réelle
- Sélection de modèles et de paramètres
- Exemples de mise en oeuvre



Prédiction

Soit un ensemble d'apprentissage $\{\mathbf{x}_i, y_i\}_{i=1}^n$ composé d'observations $\mathbf{x}_i \in \mathbb{R}^d$ de dimension d et de valeurs à prédire $y_i \in \mathcal{Y}$.

Objectif

- ▶ On cherche à apprendre à partir des données d'apprentissage une fonction de prédiction $f(\cdot) : \mathbb{R}^d \rightarrow \mathcal{Y}$.
- ▶ Types de prédiction :
 - ▶ **Classification**
 $f(\cdot)$ prédit une classe (sortie discrète) soit en classification binaire $\mathcal{Y} = \{-1, 1\}$ soit multiclasse $\mathcal{Y} = \{1, \dots, m\}$.
 - ▶ **Régression**
 $f(\cdot)$ prédit une valeur réelle ($\mathcal{Y} = \mathbb{R}$).

Fonction linéaire

$$f(\mathbf{x}) = \sum_{j=1}^d w_j x_j + b = \mathbf{w}^\top \mathbf{x} + b$$

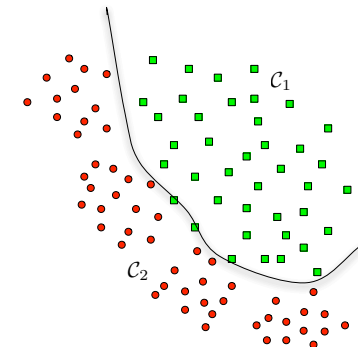
paramétrée par $\mathbf{w} \in \mathbb{R}^d$ et $b \in \mathbb{R}$



Classification binaire

Objectif

- ▶ Apprendre une fonction qui prédit la classe -1 ou 1.
- ▶ $\{\mathbf{x}_i, y_i\}_{i=1}^n \Rightarrow f(\mathbf{x})$.
- ▶ Prédiction : signe de $f(\cdot)$
- ▶ $f(\mathbf{x}) = 0$: frontière de décision.
- ▶ Paramètres :
 - ▶ Type de fonction.
 - ▶ Mesure de performance.



Méthodes

- ▶ Méthodes bayésiennes.
- ▶ Séparateur linéaire discriminant
- ▶ Séparateur à Vaste Marge.
- ▶ Arbre de décision

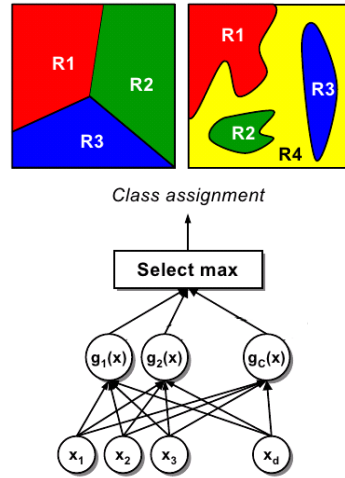
Exemples

- ▶ Reconnaissance de caractères.
- ▶ Aide au diagnostique.
- ▶ Inspection de pièces.
- ▶ Météo (pluie)



Classification multiclasse

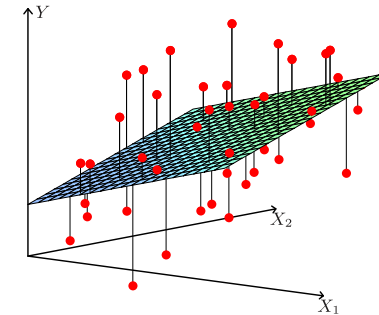
- ▶ Le rôle d'un classifieur est de **partitionner** l'espace des caractéristiques en plusieurs régions auxquels sont assignés des classes
 - ▶ les frontières s'appellent des **frontières de décision**
 - ▶ la discrimination d'un vecteur de caractéristiques \mathbf{x} consiste à déterminer à quelle région il appartient et lui assigner le label de la région
- ▶ Le classifieur peut être représenté par un ensemble de fonctions discriminantes : le classifieur affecte \mathbf{x} à la classe j si $g_j(x) > g_i(x)$ pour tout $i \neq j$



Régression

Objectif

- ▶ Apprendre une fonction qui prédit une valeur réelle.
- ▶ $\{\mathbf{x}_i, y_i\}_{i=1}^n \Rightarrow f(\mathbf{x})$.
- ▶ Paramètres :
 - ▶ Type de fonction.
 - ▶ Mesure de performance.
 - ▶ Erreur de prédiction.



Méthodes

- ▶ Moindres carrés.
- ▶ Régression ridge.
- ▶ Régression à noyaux.

Exemples

- ▶ Prédiction mouvement.
- ▶ Prédiction taux de cholestérol.
- ▶ Météo (température).

Section

Introduction

- Exemples de problèmes
- Types de problèmes RdF
- Définitions

Description/Exploration des données

- Clustering
- Estimation de densité de probabilité
- Réduction de dimension/ Visualisation

Prédiction

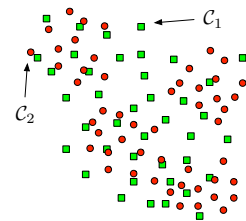
- Discrimination/Classification
- Régression

Mise en oeuvre d'un système de RdF

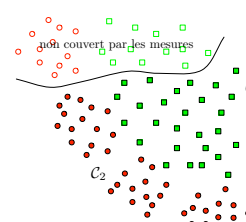
- Données réelle
- Sélection de modèles et de paramètres
- Exemples de mise en oeuvre

Données réelles (1)

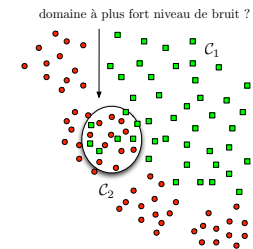
▶ Inadaptées



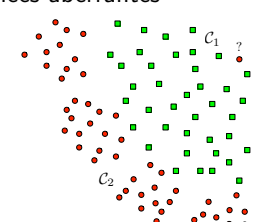
▶ Non représentative



▶ Entachées de bruit



▶ Données aberrantes

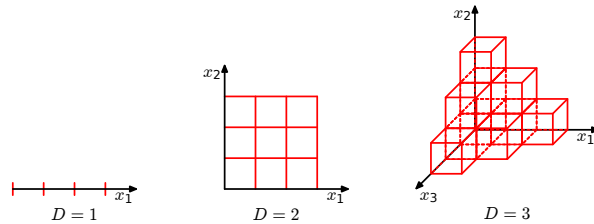


Données réelles (2)

Taille des jeux de données

On a toujours un nombre fini n de points d'apprentissage.

Malédiction de la dimensionnalité



La malédiction de la dimensionnalité exprime le fait que le nombre de données doit croître exponentiellement avec la dimension pour conserver une densité équivalente.



Sélection de modèle

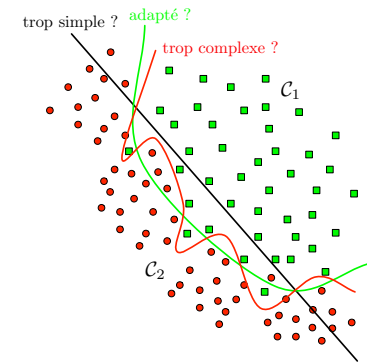
Comment sélectionner ?

Modèle	Apprentissage	Prédiction
Trop simple	--	--
Adapté	+	+
Trop complexe	++	--

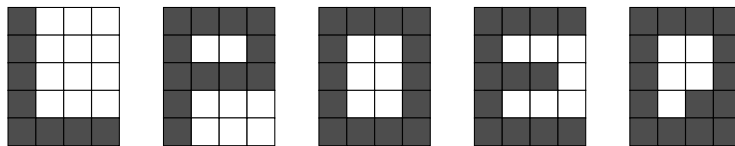
- ▶ Un modèle trop complexe provoque ce qui s'appelle du sur-apprentissage.
- ▶ On veut apprendre à prédire !

Validation

- ▶ Découpage des données en ensembles d'apprentissage/validation.
- ▶ Maximisation des performances sur les données de validation.
- ▶ La validation nécessite une bonne mesure de performances



Un exemple de tâche de reconnaissance de forme



- ▶ Développer un algorithme permettant de discriminer les lettres majuscules L, P, O, E, Q
 - ▶ Déterminer un ensemble de caractéristiques
 - ▶ Proposer un méthode de classification basé sur un arbre binaire.



Cycle de conception d'un système de RdF

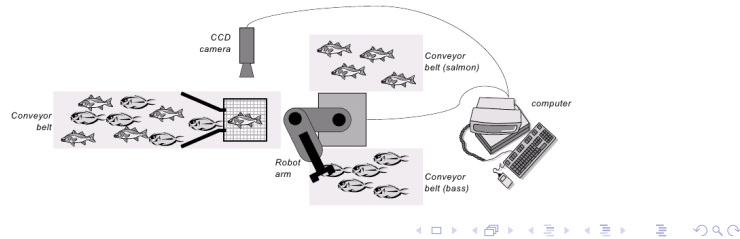
- ▶ Collecte de données
 - ▶ fastidieux et chronophage mais essentielle
 - ▶ Combien d'exemples suffisent ?
- ▶ Choix des caractéristiques
 - ▶ critique
 - ▶ peuvent être construit manuellement à partir de connaissances a priori ou automatiquement
- ▶ Choix du classifieur
 - ▶ quel modèle ?
 - ▶ comment ajuster ses paramètres ?
- ▶ Apprentissage
 - ▶ entraîner le modèle à bien "répondre" sur les données d'apprentissage
- ▶ Évaluation
 - ▶ est ce que mon modèle est bon ?
 - ▶ dilemme sur-apprentissage vs généralisation



Cycle de conception d'un système de RdF (2)

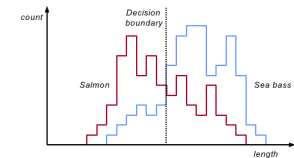
Scénario

- ▶ Une poissonnerie cherche à mettre au point un système de vision permettant de faire le tri automatique de poissons en fonctions de leurs types (saumon ou bar).
- ▶ le système est composé
 - ▶ Un tapis roulant permettant de convoier les poissons
 - ▶ 2 tapis roulant permettant de convoier les deux espèces de poisson
 - ▶ un bras robotisé permettant de faire le tri
 - ▶ un système de vision
 - ▶ un ordinateur permettant d'analyser les images et de contrôler le bras robotisé en fonction de la décision.



Cycle de conception d'un système de RdF (3)

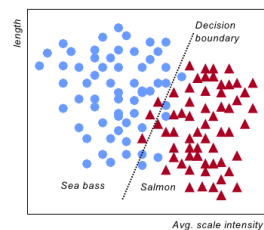
- ▶ Capteur
 - ▶ le système de vision capture une image d'un poisson arrivant sur le système de tri
- ▶ traitement d'images
 - ▶ ajustement des niveaux de gris
 - ▶ segmentation pour séparer le poisson du fond de l'image
- ▶ extraction de caractéristiques
 - ▶ On suppose qu'en moyenne, le bar est plus long qu'un saumon.
 - ▶ à partir de l'image segmentée, on estime la longueur du poisson
- ▶ Discrimination
 - ▶ Recueillir des spécimens de poissons des deux classes
 - ▶ tracer des histogrammes de longueurs pour les deux classes
 - ▶ choisir un seuil de longueur permettant de minimiser l'erreur de discrimination
 - ▶ on obtient un score décevant de 40%
 - ▶ et maintenant ?



Cycle de conception d'un système de RdF (4)

Amélioration du système RdF

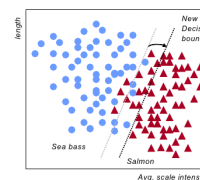
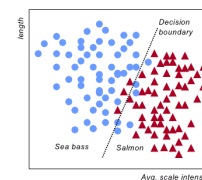
- ▶ visant un taux de reconnaissance de 95%, on essaye plusieurs caractéristiques
 - ▶ largeur, aire, position des yeux par rapport à la bouche ...
 - ▶ caractéristiques ne portant pas d'information discriminante
- ▶ finalement, on trouve une "bonne" caractéristique : le niveau de gris moyen des écailles.
- ▶ on combine "longueur" et "niveau de gris" pour améliorer la séparabilité des classes
- ▶ on calcule une fonction de décision linéaire permettant de séparer les deux classes et obtenir un taux de reconnaissance de 95.7%



Cycle de conception d'un système de RdF (5)

Coût et taux de reconnaissance

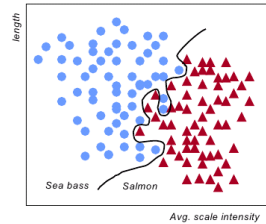
- ▶ Notre classifieur a été construit de sorte à minimiser l'erreur de discrimination
- ▶ Est ce que c'est le meilleur choix pour notre poissonnerie ?
 - ▶ le **coût** de classer un saumon comme étant un bar est que le client final trouve un "bon" goût de saumon alors qu'il a acheté un bar
 - ▶ le **coût** de classer un bar comme étant un saumon est celui du client mécontent d'avoir acheté du bar au prix du saumon
 - ▶ les coût de mauvaise classification peuvent être différent
- ▶ Intuitivement, on aimerait prendre ce coût en compte lorsqu'on construit notre frontière de décision.



Cycle de conception d'un système de RdF (6)

Généralisation

- ▶ Notre système remplit le cahier des charges avec un pourcentage de reconnaissance des exemples de 95.7%.
- ▶ En améliorant encore le système par l'utilisation d'une méthode permettant une fonction de décision non-linéaire, on aboutit à un taux de 99.9975%, avec la fonction de décision suivante



- ▶ satisfait, nous déployons notre système dans l'usine de traitement. Mais quelques semaines, le responsable de l'usine nous rappelle pour signifier qu'en pratique, le système ne reconnaît correctement que 75% des poissons
- ▶ où s'est on trompé ?