

Classification

La *classification* est la tâche consistant à **répartir** des données dans un ensemble de classes.

- « Détecter les **spams** dans un ensemble d'**emails** »
- « Repérer dans un ensemble de **tweets** ceux qui sont **sarcastiques** ou **ironiques** »
- « Classer un ensemble de **textes** par **genre littéraire** »

```
graph LR; Donnée([Donnée]) --> Programme[Programme de classification]; Programme --> Classe([Classe]); ClassesPossibles([Classes possibles]) --> Programme
```

5

Classification

La *classification* est la tâche consistant à **répartir** des données dans un ensemble de classes.

- « Détecter les **spams** dans un ensemble d'**emails** »
- « Repérer dans un ensemble de **tweets** ceux qui sont **sarcastiques** ou **ironiques** »
- « Classer un ensemble de **textes** par **genre littéraire** »

```
graph LR; Donnée([Donnée]) --> Programme[Programme de classification]; Programme --> Classe([Classe]); ClassesPossibles([Classes possibles]) --> Programme
```

- Les **classes** sont connues à l'avance
- Chaque **donnée** appartient à une et une seule **classe**.

5

Classification

Classification par règles expertes

Classification par règles expertes

La façon traditionnelle de faire de la classification

- Faire écrire par des experts du domaine un ensemble de règles associant un exemple à une classe

6

Classification par règles expertes

La façon traditionnelle de faire de la classification

- Faire écrire par des experts du domaine un ensemble de règles associant un exemple à une classe
- Exemple : classification du sentiment
 - Établir un lexique affectif mot → ±1
 - Associer à chaque texte la somme des polarités de ses termes
 - Un texte est positif si la somme est positive, négatif sinon

6

Classification par règles expertes

La façon traditionnelle de faire de la classification

- Faire écrire par des experts du domaine un ensemble de règles associant un exemple à une classe
- Exemple : classification du sentiment
 - Établir un lexique affectif mot → ±1
 - Associer à chaque texte la somme des polarités de ses termes
 - Un texte est positif si la somme est positive, négatif sinon
- Qu'en pensez-vous?

6

Classification par règles expertes

La façon traditionnelle de faire de la classification

- Faire écrire par des experts du domaine un ensemble de règles associant un exemple à une classe
- Exemple : classification du sentiment
 - Établir un lexique affectif mot → ±1
 - Associer à chaque texte la somme des polarités de ses termes
 - Un texte est positif si la somme est positive, négatif sinon
- Qu'en pensez-vous?
- Pour notre exemple de classification culture/société, comment faire?

6

Classification par apprentissage artificiel

Rappel
Un **programme d'apprentissage artificiel** est un programme générant à partir d'**exemples** traités par des humains un **modèle** qui sert de ressource à un **programme réalisant une tâche**.

```
graph LR; Exemples([Exemples]) --> ProgrammeApprentissage[Programme d'apprentissage]; ProgrammeApprentissage --> Modèle([Modèle]); Entrée([Entrée]) --> ProgrammeTache[Programme réalisant la tâche]; ProgrammeTache --> Résultat([Résultat]); Modèle --> ProgrammeTache
```

Schéma d'une approche par apprentissage artificiel

7

Exemple : régression

« Sachant que ma maison fait $s\text{ m}^2$, en fonction du marché, quel prix puis-je en attendre? »

8

«Sachant que ma maison fait $s\text{ m}^2$, en fonction du marché, quel prix puis-je en attendre?»

Hypothèse Il existe une fonction f qui à la surface x d'une maison associe son prix de vente

«Sachant que ma maison fait $s\text{ m}^2$, en fonction du marché, quel prix puis-je en attendre?»

Hypothèse Il existe une fonction f qui à la surface x d'une maison associe son prix de vente

Données Le prix de vente de 8 maisons et leurs surfaces respectives

«Sachant que ma maison fait $s\text{ m}^2$, en fonction du marché, quel prix puis-je en attendre?»

Hypothèse Il existe une fonction f qui à la surface x d'une maison associe son prix de vente

Données Le prix de vente de 8 maisons et leurs surfaces respectives

Objectif Déterminer l'expression de f et en déduire $f(s)$, le prix de vente de ma maison

→ Problème : f pourrait avoir n'importe quelle forme!

Choix d'une famille de fonctions

Problème : f pourrait avoir n'importe quelle forme!

Choix d'une famille de fonctions

Problème : f pourrait avoir n'importe quelle forme!

- On choisit une famille de fonctions et on y cherche celle qui semble le mieux correspondre à nos exemples
- Par exemple avec des polynômes :

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

Choix d'une famille de fonctions

Problème : f pourrait avoir n'importe quelle forme!

- On choisit une famille de fonctions et on y cherche celle qui semble le mieux correspondre à nos exemples
- Par exemple avec des polynômes :

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

- En général «correspondre» au sens des moindres carrés

Choix d'une famille de fonctions

Problème : f pourrait avoir n'importe quelle forme!

- On choisit une famille de fonctions et on y cherche celle qui semble le mieux correspondre à nos exemples
- Par exemple avec des polynômes :

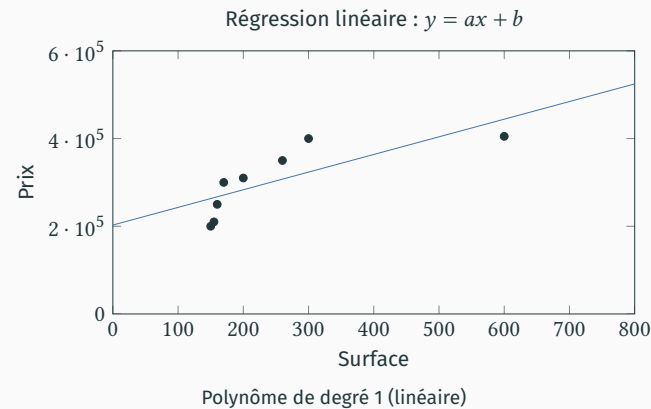
$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

- En général «correspondre» au sens des moindres carrés

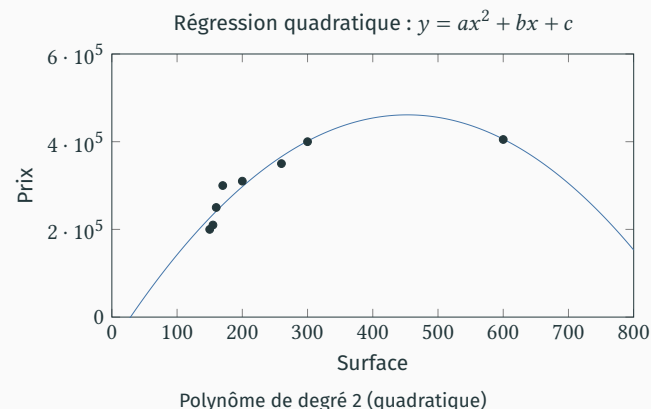
Vocabulaire Le degré (fixé manuellement) est un **hyperparamètre**, les coefficients a_i appris automatiquement sont des **paramètres**. Déterminer les paramètres qui minimisent le coût (ici la distance entre le modèle et les exemple) est une tâche d'**optimisation**.

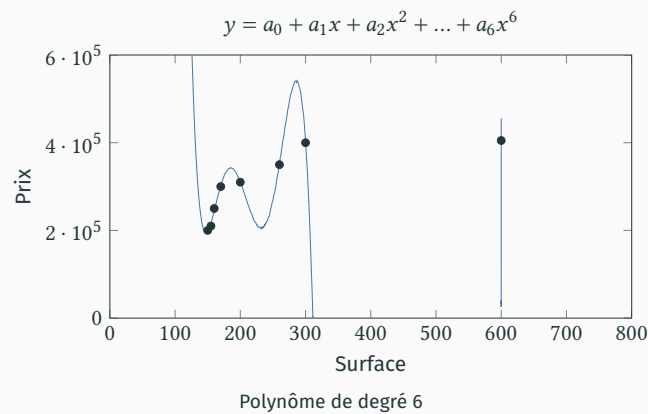
Étant donné une surface, on peut alors donner le prix correspondant : on donc a appris à prédire les prix!

Régression linéaire



Régression quadratique





La figure 4 est un exemple de cas de **surapprentissage** (*overfitting*)

- Avoir choisi le degré 6 permet de passer exactement par tous les points d'entraînement
- Mais le modèle est très mauvais!

La figure 4 est un exemple de cas de **surapprentissage** (*overfitting*)

- Avoir choisi le degré 6 permet de passer exactement par tous les points d'entraînement
- Mais le modèle est très mauvais!

Comment trouver un modèle *vraiment* bon ?

- En limitant le nombre de paramètres
- En limitant la précision des paramètres
- Par le choix des hyperparamètres
- En utilisant un ensemble de développement

« Apprentissage » ?

Ce qu'on appelle *apprendre* pour une machine consiste à transformer des exemples (x_i, y_i) en règle $f : x \mapsto y$. Soit

- Étant donné un espace de recherche F
- Étant donnée une mesure de qualité q sur F
- Trouver $f \in F$ qui maximise q

« Apprentissage » ?

Ce qu'on appelle *apprendre* pour une machine consiste à transformer des exemples (x_i, y_i) en règle $f : x \mapsto y$. Soit

- Étant donné un espace de recherche F
- Étant donnée une mesure de qualité q sur F
- Trouver $f \in F$ qui maximise q

En général F est paramétré par des nombres a_1, \dots, a_n et on se ramène ainsi à chercher leurs valeurs pour que $q(f_{a_1, \dots, a_n})$ soit maximal.

« Automatique » ?

Il reste encore à l'humain à

1. Choisir un espace de recherche adapté
2. Choisir une mesure de qualité adaptée
3. Choisir une procédure d'optimisation adaptée

En tenant compte

Pour 1,

- De l'expressivité du modèle et de son adéquation au problème
- De la lisibilité du résultat

« Automatique » ?

Il reste encore à l'humain à

1. Choisir un espace de recherche adapté
2. Choisir une mesure de qualité adaptée
3. Choisir une procédure d'optimisation adaptée

En tenant compte

Pour 2,

- De l'adéquation aux données d'apprentissage
- La généralisabilité (éviter le surapprentissage)
- La simplicité : rasoir d'Ockham

« Automatique » ?

Il reste encore à l'humain à

1. Choisir un espace de recherche adapté
2. Choisir une mesure de qualité adaptée
3. Choisir une procédure d'optimisation adaptée

En tenant compte

Pour 3,

- De l'incrémentalité
- Du temps de calcul

On verra dans la suite que ces trois choix sont souvent dépendants.

« Automatique » ?

Il reste encore à l'humain à

1. Choisir un espace de recherche adapté
2. Choisir une mesure de qualité adaptée
3. Choisir une procédure d'optimisation adaptée

- Il n'y a pas de recette miracle qui marche à chaque fois : « *There is no free lunch* »
- Pour cette raison, on parle aussi d'apprentissage *artificiel* plutôt qu'*automatique* cf *Machine Learning*

Plan des réjouissances	Think fast	
<p>On présente dans la suite plusieurs techniques classiques de classification par apprentissage artificiel suivant avec pour chacune</p> <ul style="list-style-type: none"> Espace de recherche Hyperparamètres Technique de recherche Propriétés et usages 	<p>Dans un problème de classification, dans l'ensemble d'entraînement, il y a 1798 exemples de la classe A et 2 de la classe B.</p> <p>On vous donne un nouvel exemple dont vous ne connaissez pas la classe sans plus d'information. Dans quelle classe le rangez-vous ?</p>	<div>Classification</div> <div>Algorithme de la classe majoritaire</div>
16	17	
Algorithme de la classe majoritaire	Algorithme de la classe majoritaire	Algorithme de la classe majoritaire
<p>Dans Weka classifieurs > rules > ZeroR</p> <p>Espace de recherche Ensemble des fonctions constantes</p> <p>Autrement dit : on renvoie toujours la même classe quelle que soit l'entrée.</p>	<p>Dans Weka classifieurs > rules > ZeroR</p> <p>Espace de recherche Ensemble des fonctions constantes</p> <p>Autrement dit : on renvoie toujours la même classe quelle que soit l'entrée.</p> <p>Hyperparamètres Aucun !</p>	<p>Dans Weka classifieurs > rules > ZeroR</p> <p>Espace de recherche Ensemble des fonctions constantes</p> <p>Autrement dit : on renvoie toujours la même classe quelle que soit l'entrée.</p> <p>Hyperparamètres Aucun !</p>
18	18	18
Algorithme de la classe majoritaire	Propriétés	
<p>Dans Weka classifieurs > rules > ZeroR</p> <p>Espace de recherche Ensemble des fonctions constantes</p> <p>Autrement dit : on renvoie toujours la même classe quelle que soit l'entrée.</p> <p>Hyperparamètres Aucun !</p> <p>Technique de recherche On choisit la fonction qui renvoie la classe majoritaire dans l'ensemble d'apprentissage.</p>	<ul style="list-style-type: none"> Très rapide en apprentissage comme en test Très peu performant <ul style="list-style-type: none"> Dans le meilleur des cas (si les classes sont déséquilibrées) on fait un peu mieux que le hasard <p>Il s'agit plutôt d'une <i>baseline</i> : un système sérieux devrait en principe faire mieux.</p>	<div>Classification</div> <div><i>k</i>-plus proches voisins</div>
18	19	

Dans Weka classifiers > lazy > IBk

Espace de recherche Le singleton $\{f\}$, où f est la fonction qui à un exemple x associe la classe majoritaire dans l'ensemble des k données d'entraînement les plus proches de x .

En général on utilise la distance euclidienne ou le coefficient d'overlap (pour des vecteurs booléens)

Hyperparamètres k , un entier

- Plus petit que le nombre d'exemples d'apprentissages
- Idéalement non-multiple du nombre de classes

Dans Weka classifiers > lazy > IBk

Espace de recherche Le singleton $\{f\}$, où f est la fonction qui à un exemple x associe la classe majoritaire dans l'ensemble des k données d'entraînement les plus proches de x .

En général on utilise la distance euclidienne ou le coefficient d'overlap (pour des vecteurs booléens)

Dans Weka classifiers > lazy > IBk

Espace de recherche Le singleton $\{f\}$, où f est la fonction qui à un exemple x associe la classe majoritaire dans l'ensemble des k données d'entraînement les plus proches de x .

En général on utilise la distance euclidienne ou le coefficient d'overlap (pour des vecteurs booléens)

Hyperparamètres k , un entier

- Plus petit que le nombre d'exemples d'apprentissages
- Idéalement non-multiple du nombre de classes

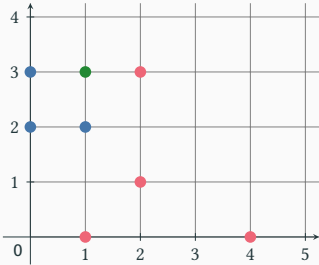
Dans Weka classifiers > lazy > IBk

Espace de recherche Le singleton $\{f\}$, où f est la fonction qui à un exemple x associe la classe majoritaire dans l'ensemble des k données d'entraînement les plus proches de x .

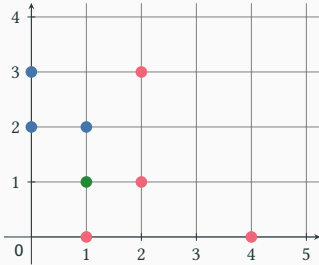
En général on utilise la distance euclidienne ou le coefficient d'overlap (pour des vecteurs booléens)

Hyperparamètres k , un entier

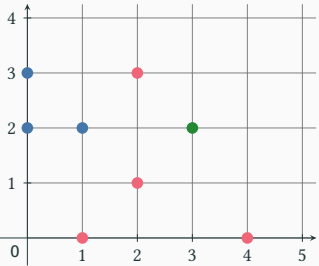
- Plus petit que le nombre d'exemples d'apprentissages
- Idéalement non-multiple du nombre de classes



Exemple de classification : k-NN



Exemple de classification : k-NN



Exemple de classification : k-NN

- Aucune recherche → apprentissage immédiat
- Coût en opération : Taille de l'ensemble d'entraînement × coût d'un calcul de distance
- Pour être efficace : ensemble d'entraînement suffisamment dense
→ D'où sa faible utilisation en pratique en fouille de texte (à part comme baseline)
- Fréquemment utilisé pour des moteurs de recommandation