

Présentation

Le b.a-ba du machine learning appliqué au risque client : une approche basée sur la marge optimale



Le b.a-ba du machine learning appliqué au client : une approche basée sur la marge optimale

Deux slides de présentation

Les notions de base du machine learning

Application au crédit client : une approche basée sur la marge

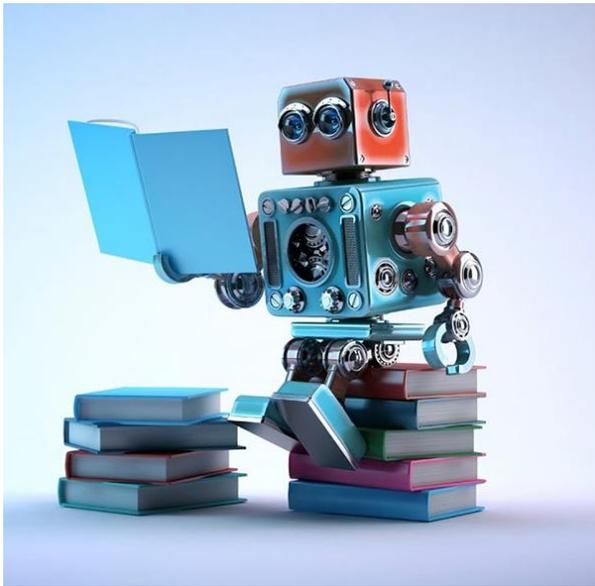
Conclusion

Blitz en quelques chiffres

- 40 000 caisses connectées
- 42 millions de transactions traitées chaque année
- Un flux traité de 3000 M€
- Une réponse en moins de 30 ms dans 90% des cas et moins de 300 ms dans 100% des cas
- Pas d'appel téléphonique en ligne de caisse
- 40 serveurs dans 2 data centers
- Régulé par le General Data Protection Regulation (EU 2016/679)
- 2,3 M€ de chiffre d'affaires

Un savoir faire basé sur 3 piliers techniques

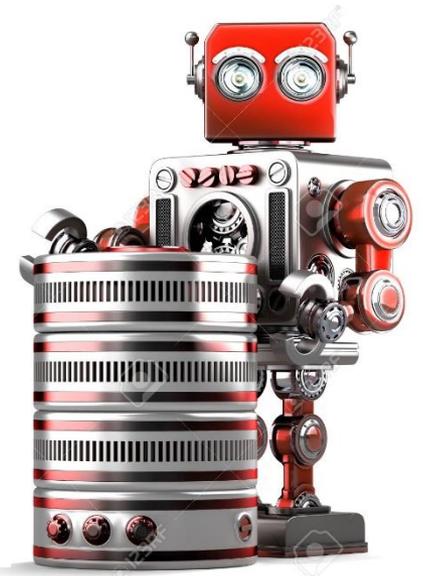
Intelligence artificielle
(Machine learning)



Informatique temps réel



Gestion de larges bases de données





Le b.a-ba du machine learning appliqué au client : une approche basée sur la marge optimale

Deux slides de présentation

Les notions de base du machine learning

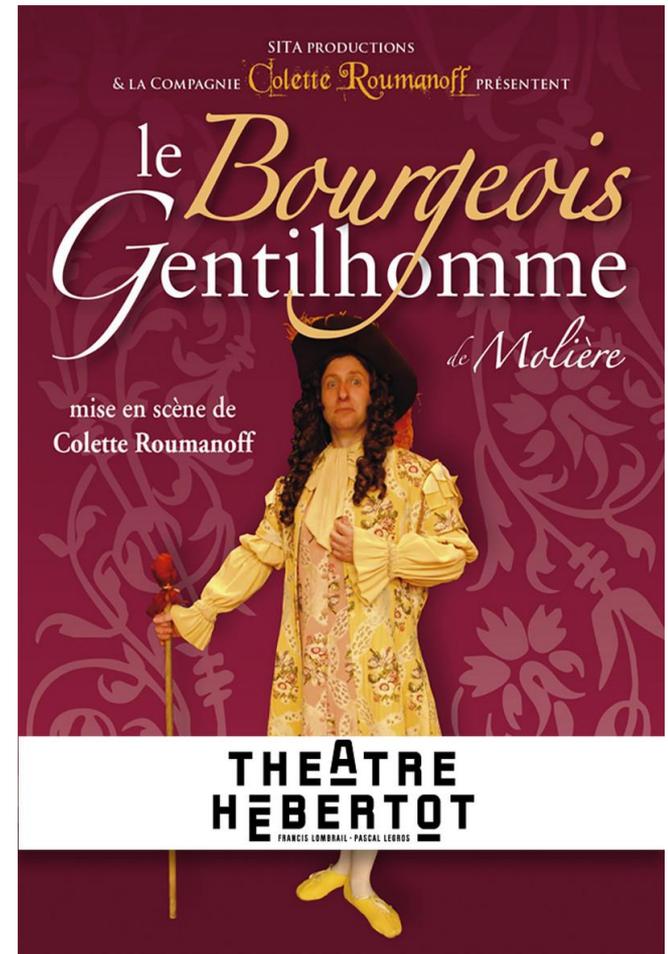
Application au crédit client : une approche basée sur la marge

Conclusion

L'objectif de cette première partie est d'un peu démystifier l'intelligence artificielle...

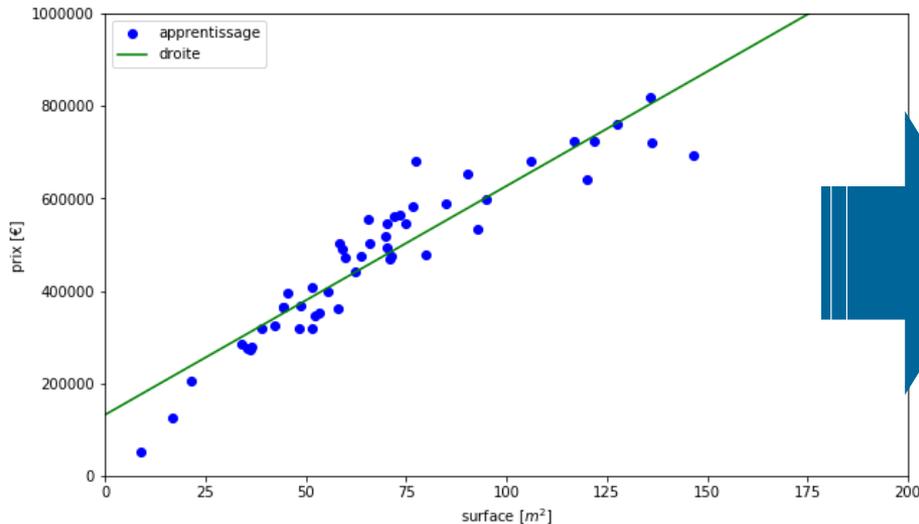
Par ma foi ! il y a plus de quarante ans que je dis de la prose sans que j'en suse rien, et je vous suis le plus obligé du monde de m'avoir appris cela.

Monsieur Jourdain



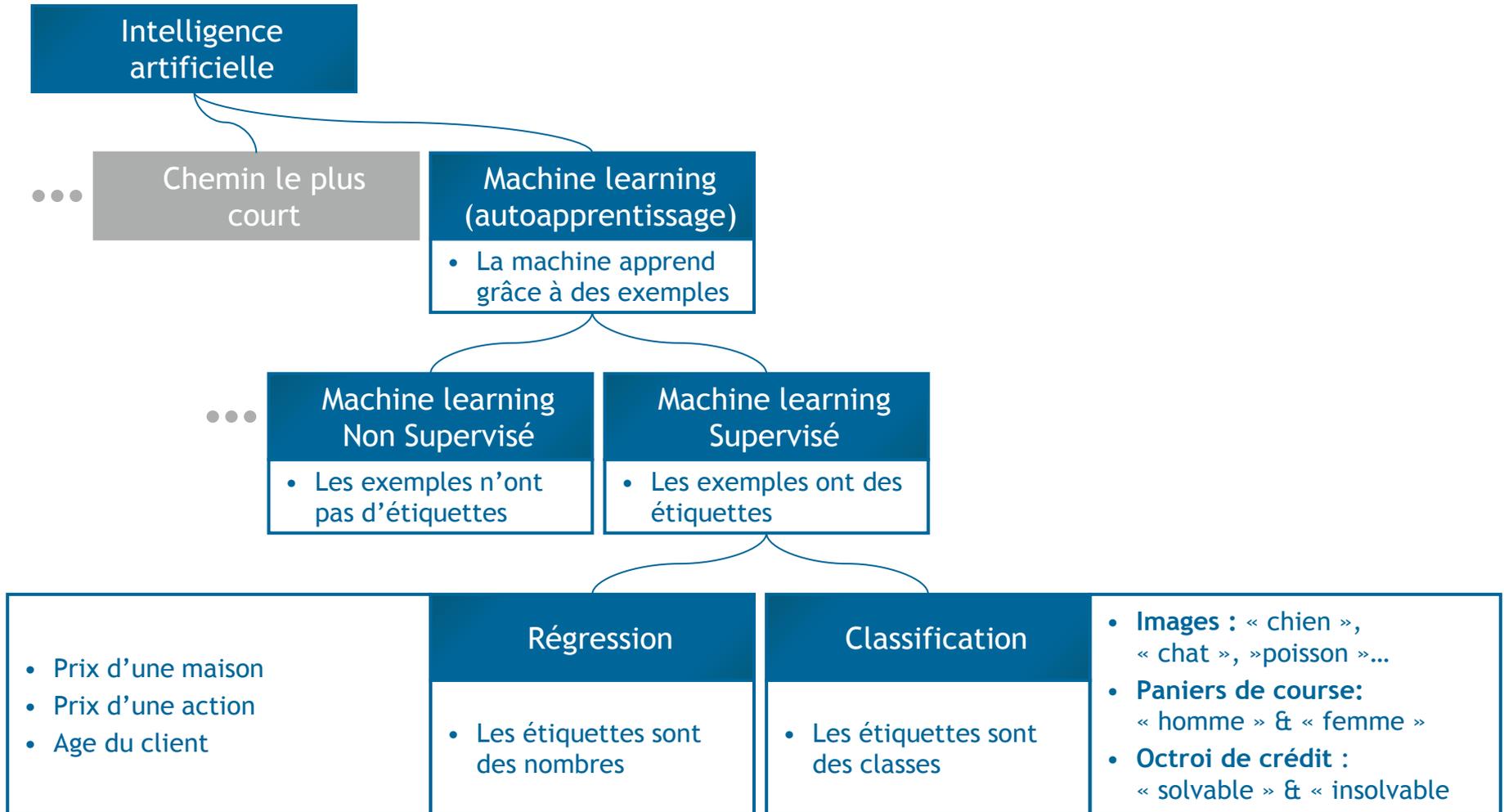
...et de vous transmettre le vocabulaire de base

L'intelligence artificielle débute avec des notions très simples...

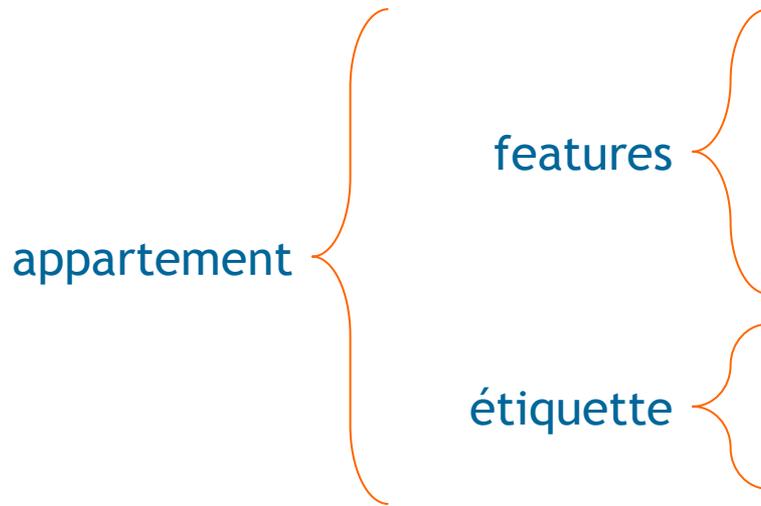


...mais la frontière actuelle repose sur des techniques très complexes

Nous allons apprendre les notions de base en machine learning supervisé



Chaque exemple est décrit par des « features »



- Surface
- Etage
- Quartier
- Garage (O/N)

- Prix

Objectif :
Apprendre avec des exemples...
Puis trouver l'étiquette à partir de la
seule connaissance des features

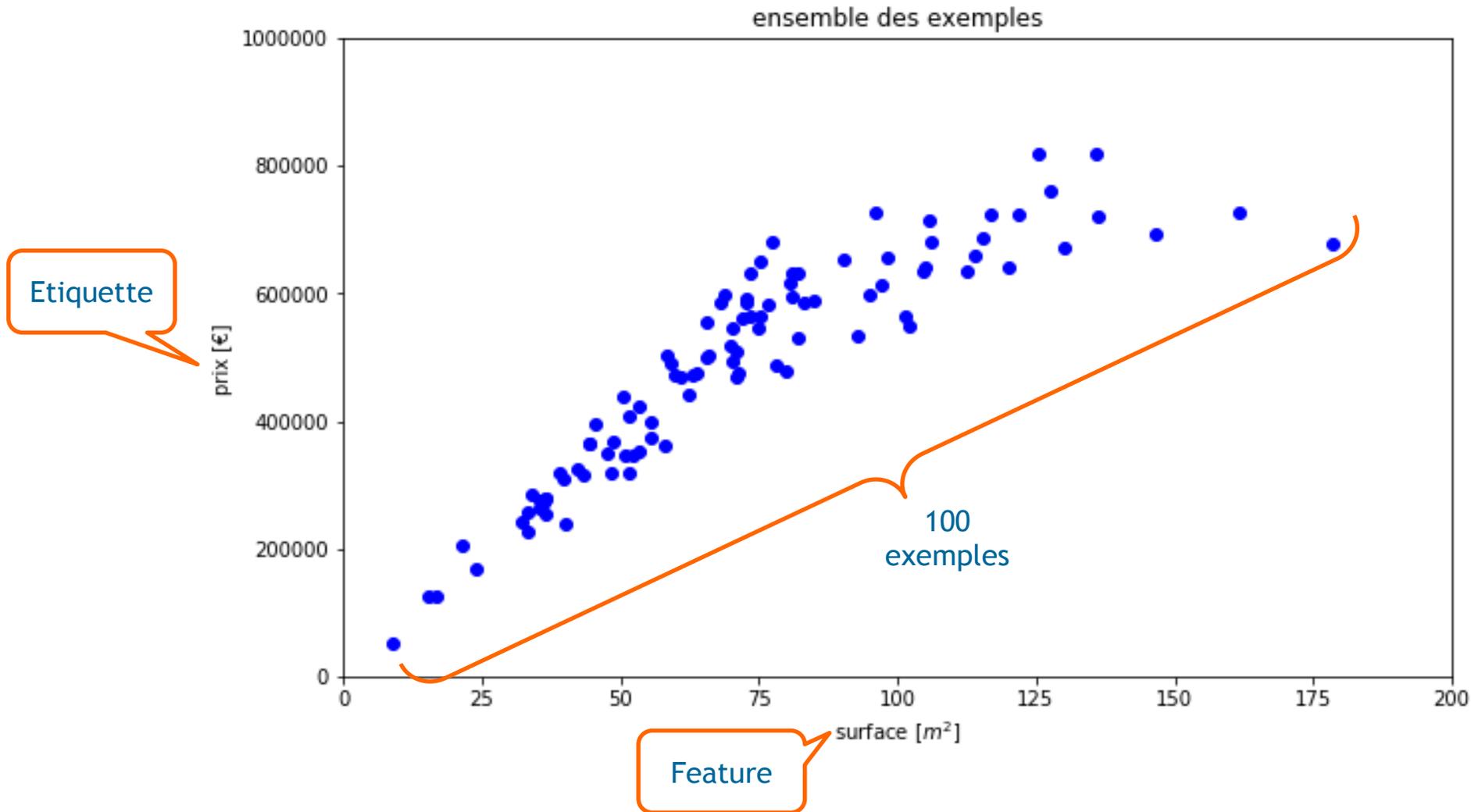




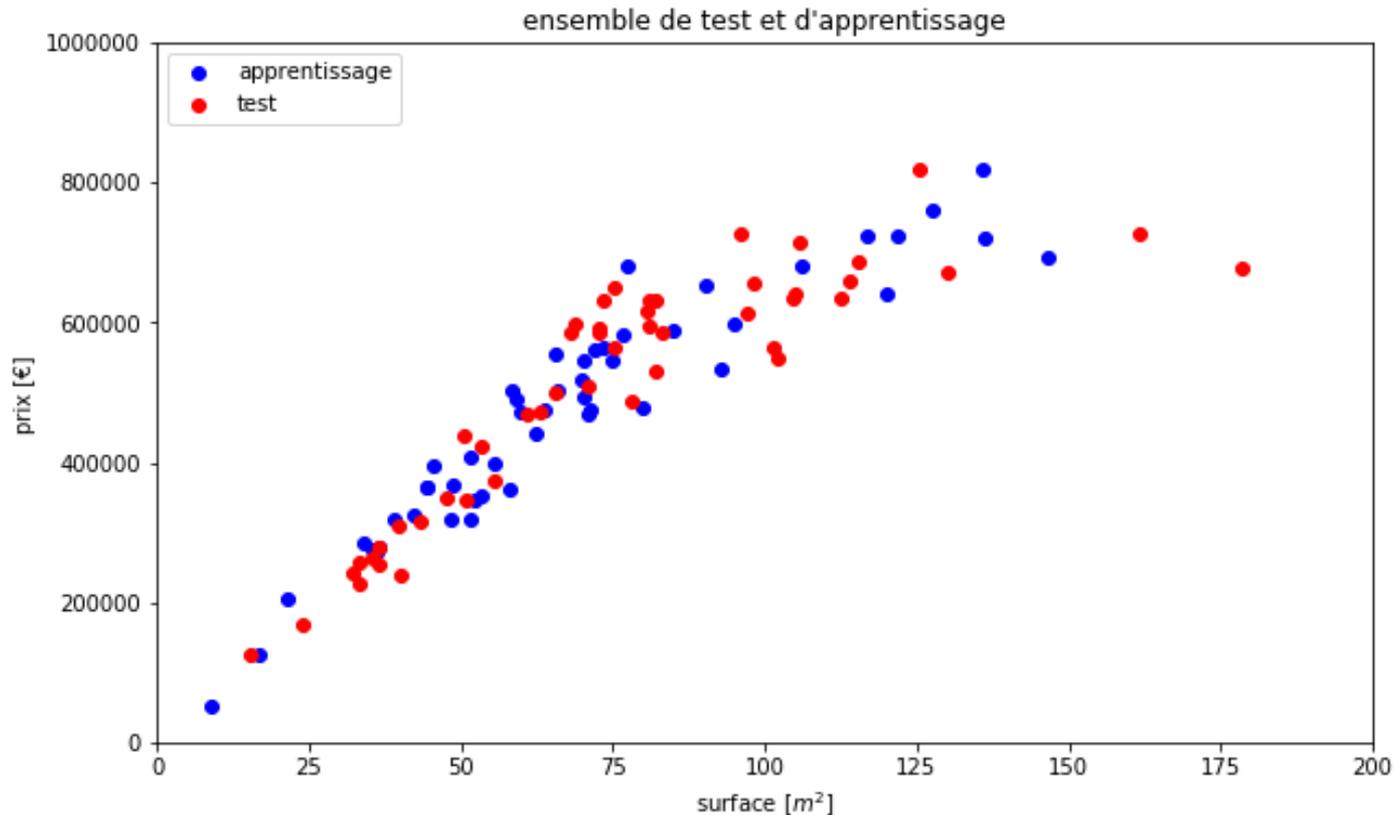
Notre liste de vocabulaire

- Machine learning ou apprentissage automatique
- Supervisé
- Non supervisé
- Exemples
- Etiquette
- Features

Nous allons programmer un robot qui estime le prix d'un appartement à partir de la connaissance de sa surface

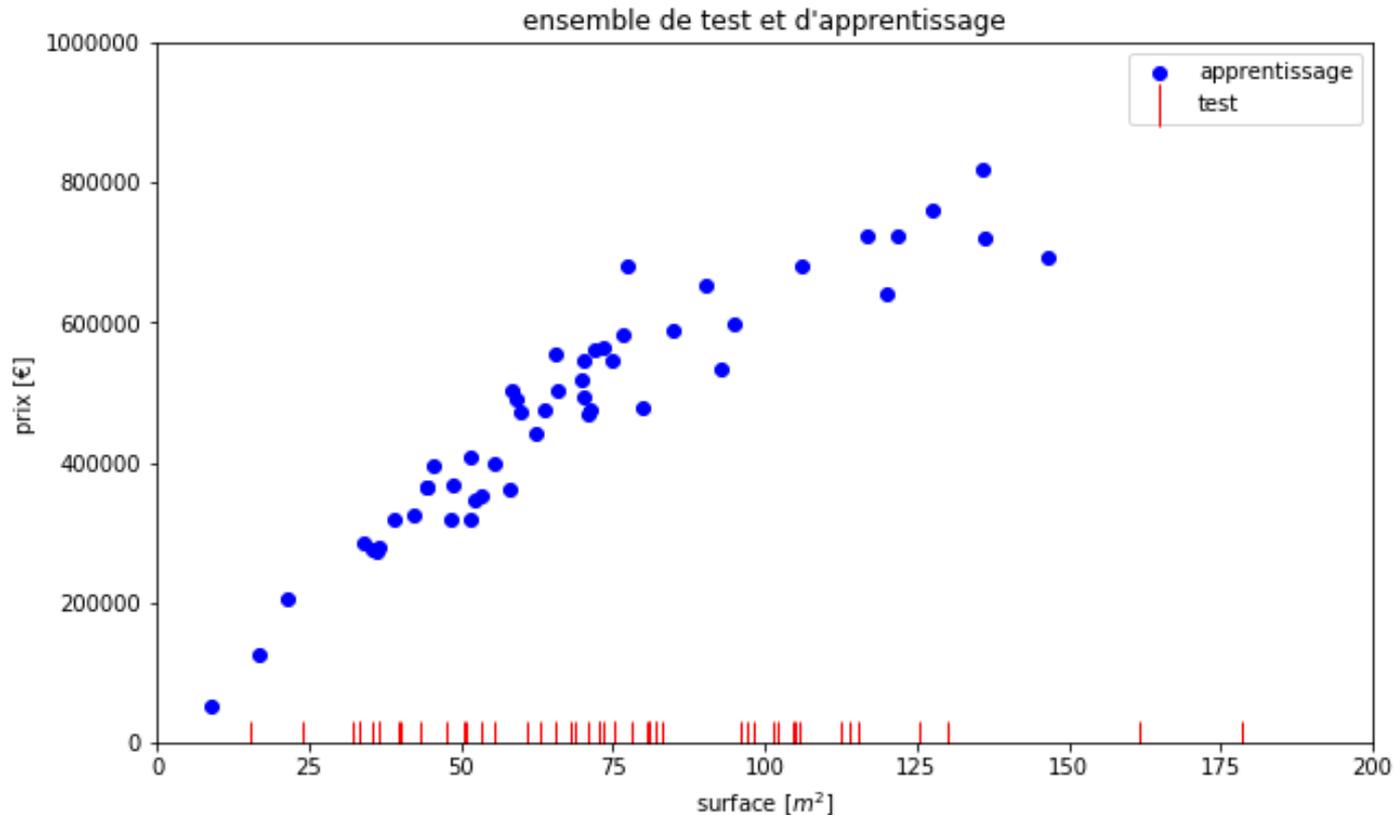


Une base d'apprentissage et une base de test sont tirées au hasard



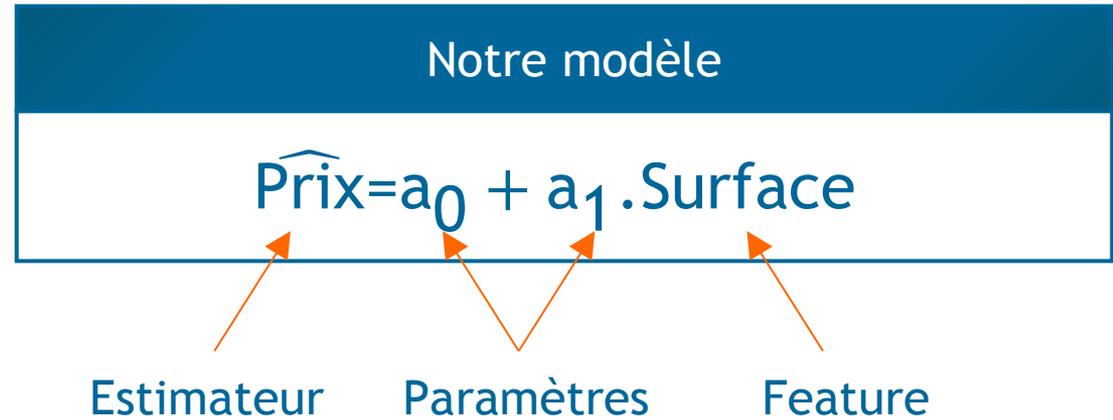
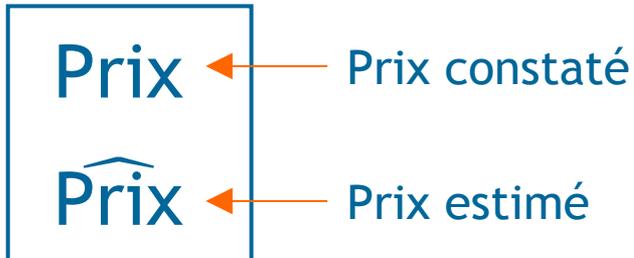
- La base d'apprentissage va servir à apprendre la fonction qui donne le prix à partir de la surface
- Le base de test va servir à vérifier que le prix estimé est proche du prix constaté

Une base d'apprentissage et une base de test sont tirées au hasard

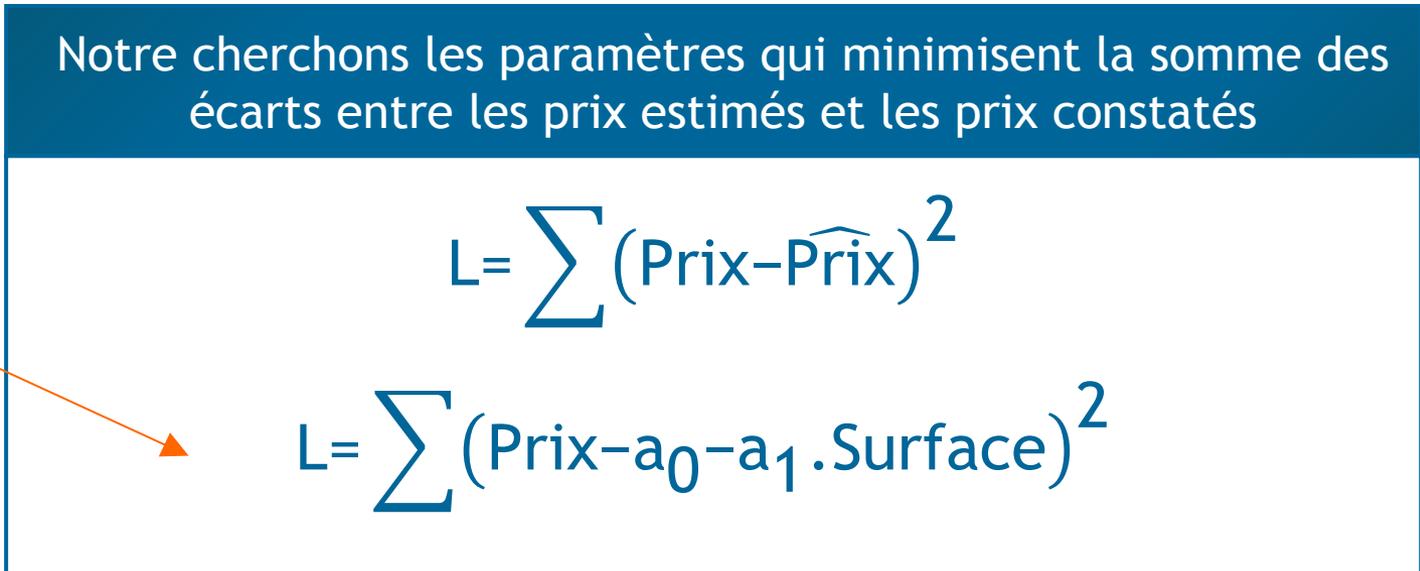
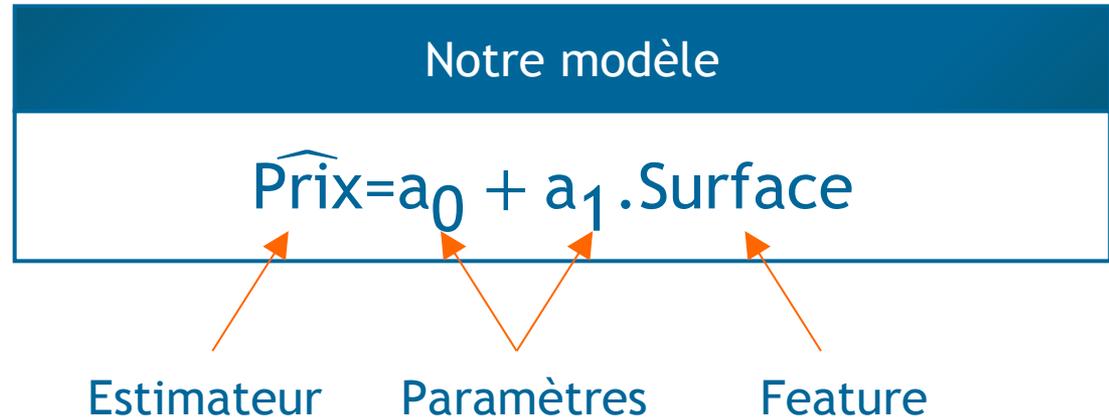
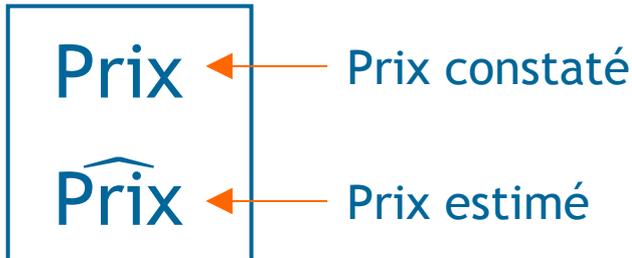


- Seule la base d'apprentissage va être utilisée pour apprendre le modèle

Commençons par un modèle linéaire

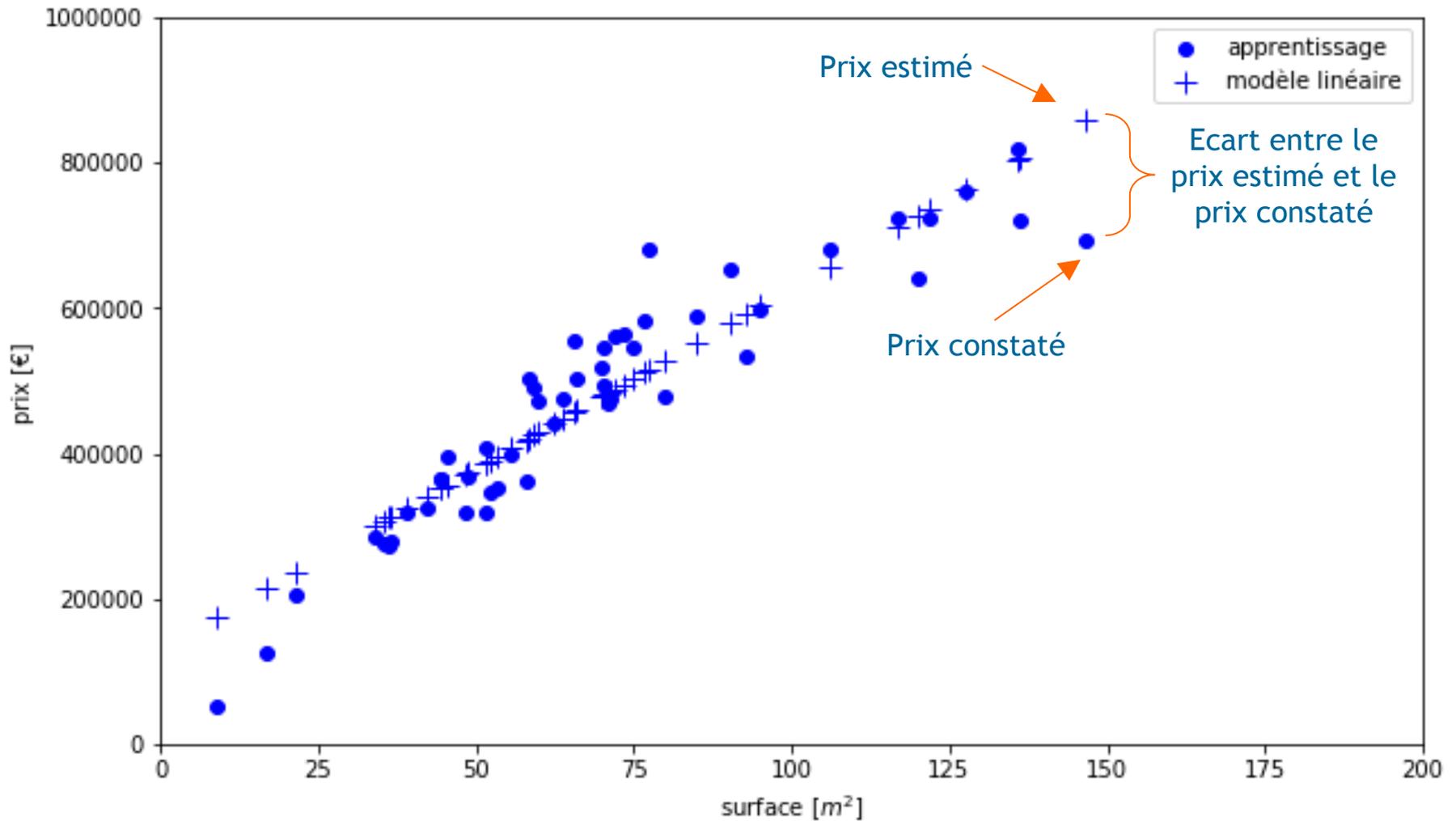


Commençons par un modèle linéaire



Fonction Loss =
fonction objectif
à minimiser

Visualisons les exemples et leurs estimateurs



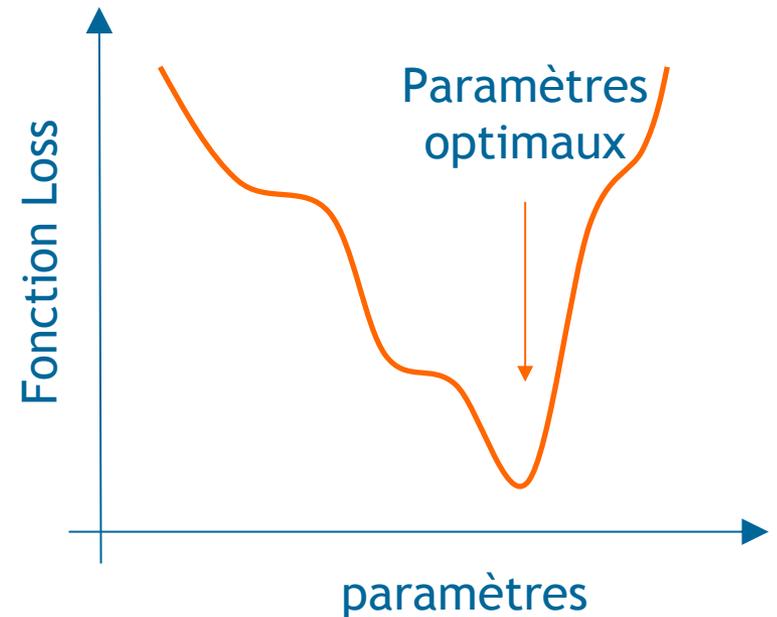
Il y a plein de choix possibles pour la fonction Loss !

Faire du machine learning c'est minimiser une fonction objectif

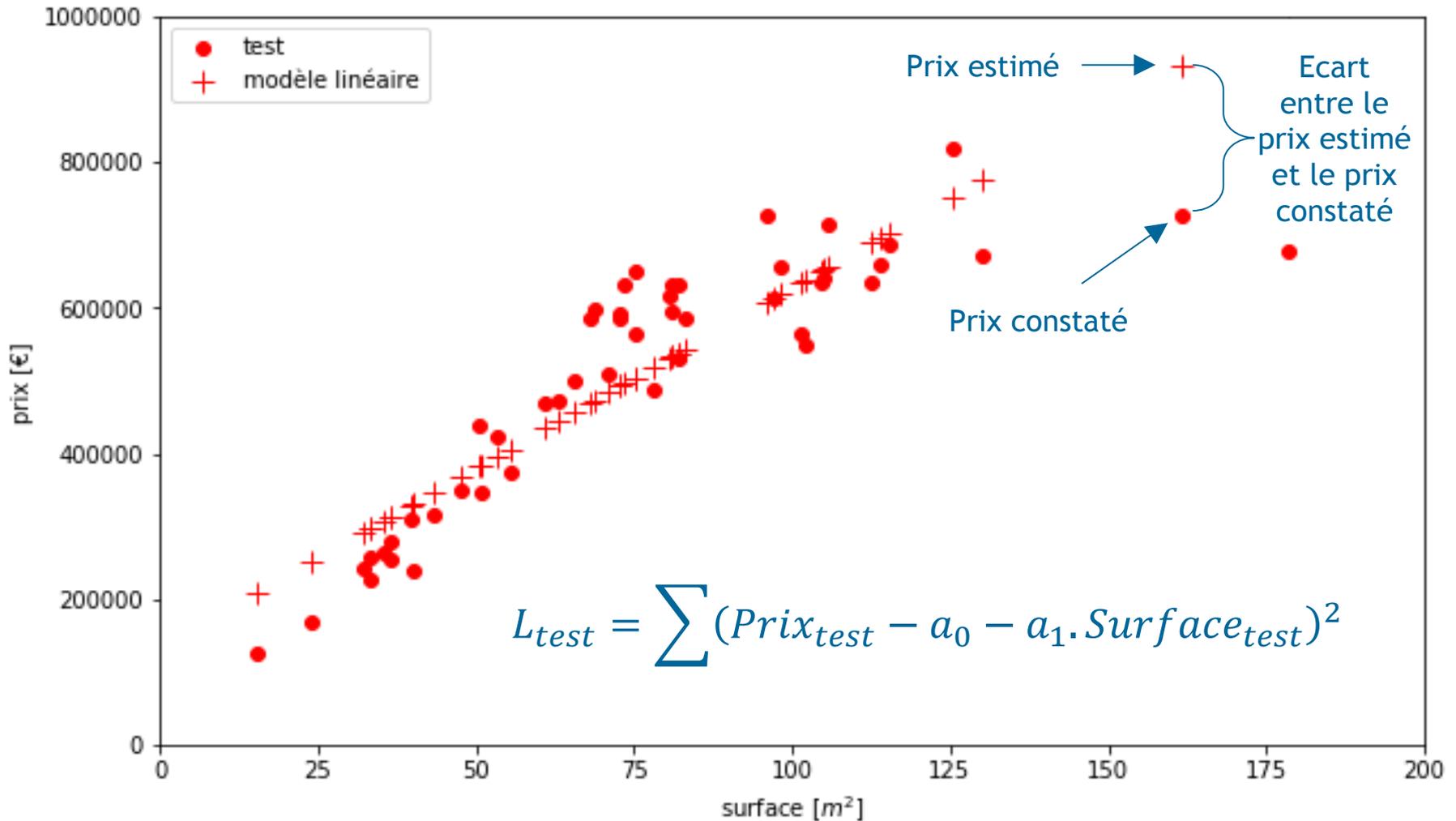
$$L = \sum (\text{Prix} - \widehat{\text{Prix}})^2$$

$$L = \sum |\text{Prix} - \widehat{\text{Prix}}|$$

$$L = \sum (\text{Prix} - \widehat{\text{Prix}})^4$$



Nous pouvons maintenant valider notre modèle sur notre base de test





Notre liste de vocabulaire

- Machine learning ou apprentissage automatique
- Supervisé
- Non supervisé
- Exemples
- Etiquette
- Features
- Base d'apprentissage
- Base de test
- Modèle
- Paramètre
- Fonction loss ou fonction objectif

Complexifions notre modèle...

Notre modèle du second degrés (degrés 2)

$$\widehat{\text{Prix}} = a_0 + a_1 \cdot \text{Surface} + a_2 \cdot \text{Surface}^2$$

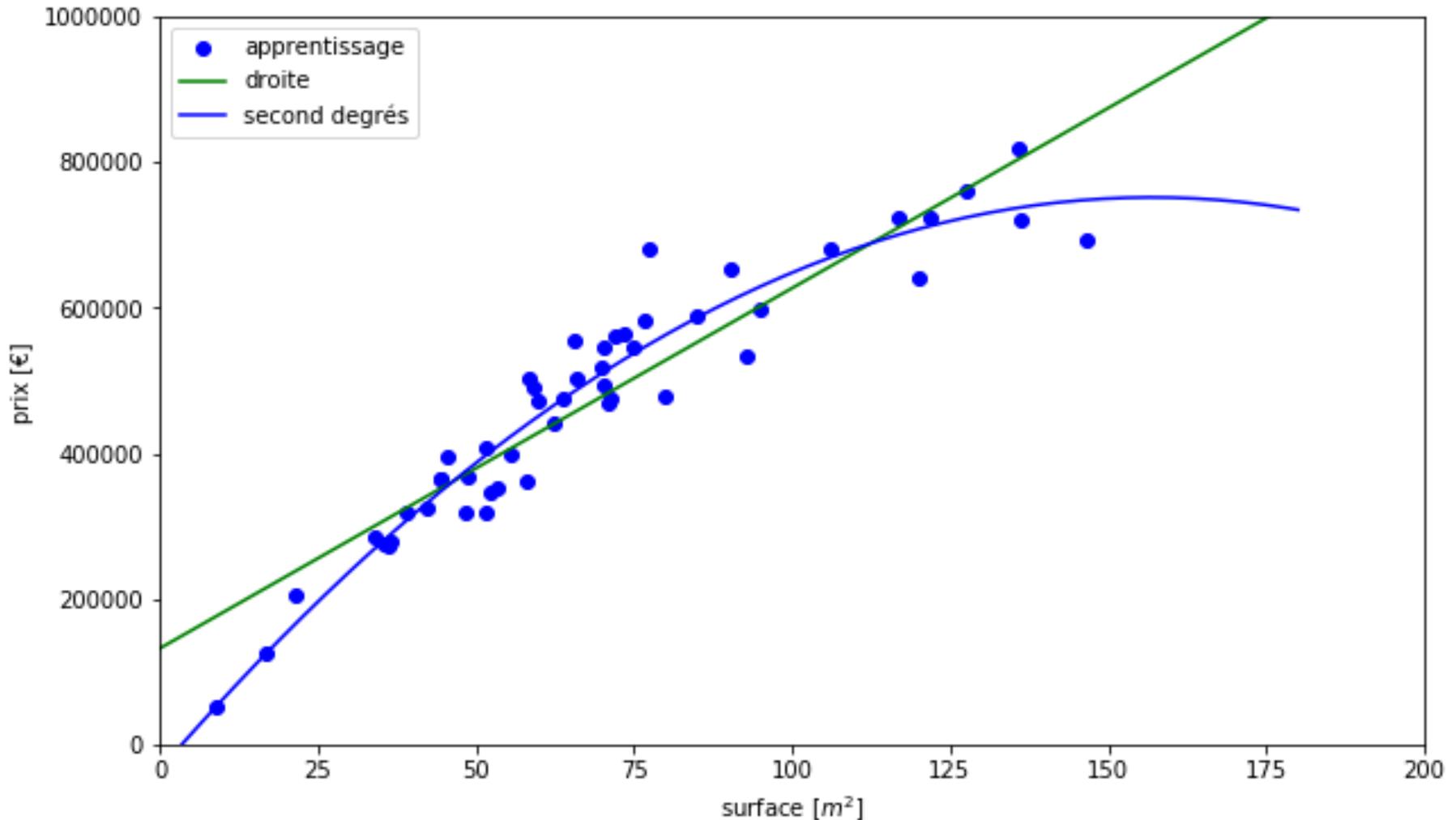
Estimateur

Degrés 2 : 3 paramètres

Nous conservons la même fonction Loss

$$L = \sum (\text{Prix} - \widehat{\text{Prix}})^2$$

Ce modèle plus complexe se rapproche mieux des exemples de la base d'apprentissage



Complexifions encore notre modèle pour se rapprocher encore plus les exemples d'apprentissage

Notre modèle du cinquième degré (degrés 5)

$$\widehat{\text{Prix}} = a_0 + a_1 \cdot \text{Surface} + a_2 \cdot \text{Surface}^2 + \dots + a_5 \cdot \text{Surface}^5$$

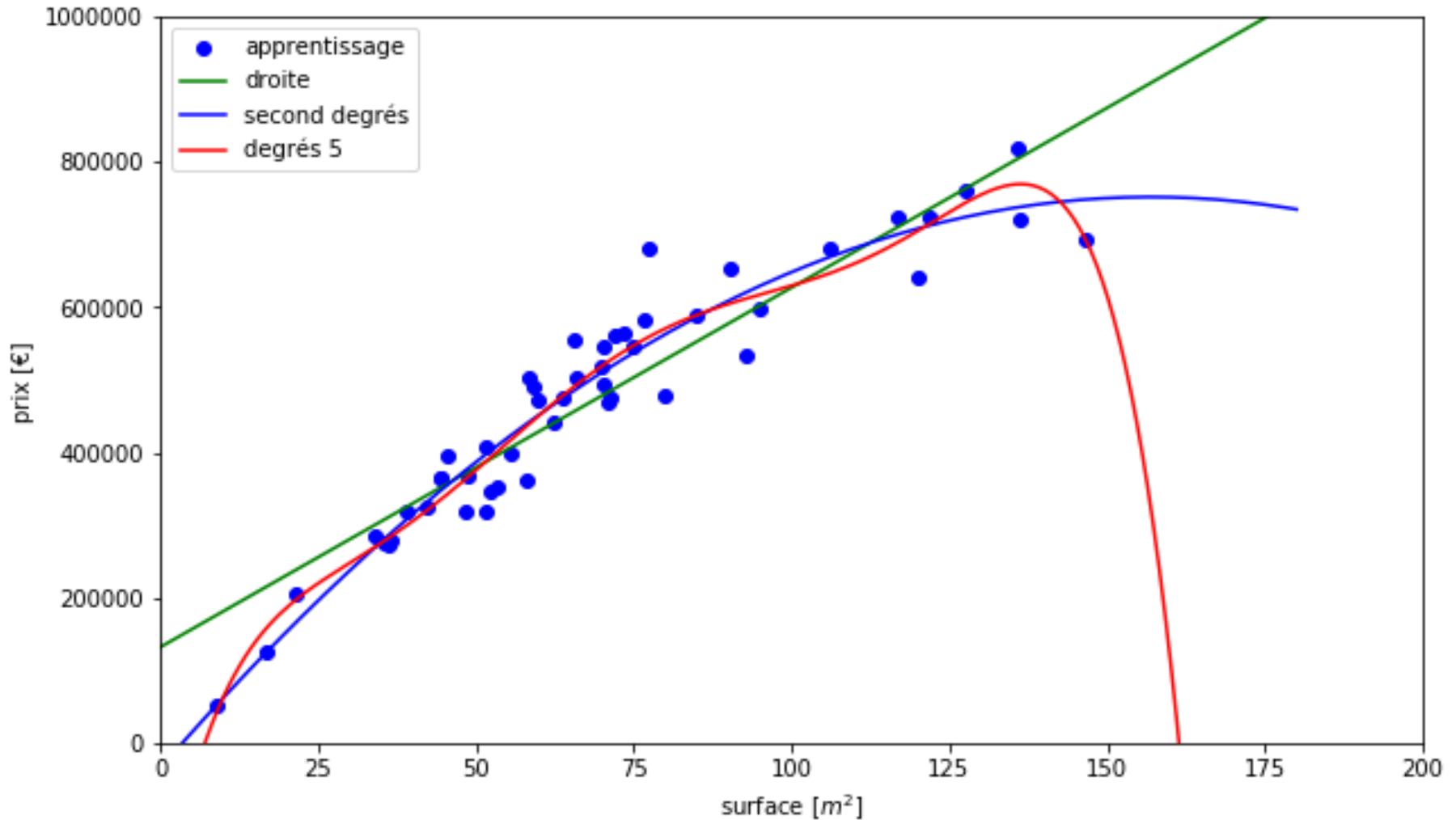
Estimateur

Degrés 5 : 6 paramètres

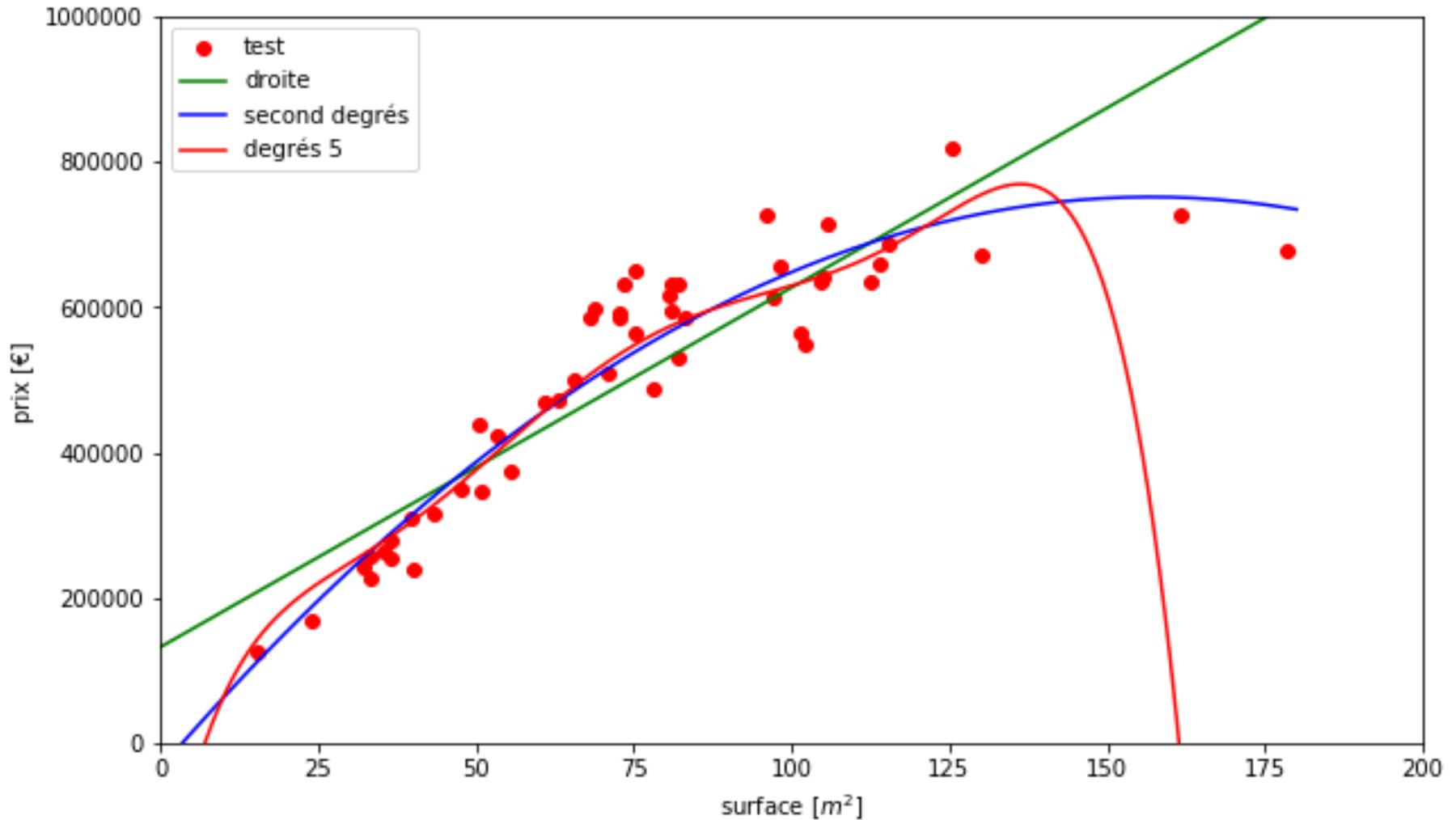
Nous conservons la même fonction Loss

$$L = \sum (\text{Prix} - \widehat{\text{Prix}})^2$$

Le modèle s'approche plus des exemples d'apprentissage...



...mais s'éloigne des exemples de test



Complexifions encore notre modèle...

Notre modèle du 25^{ième} degrés (degrés 25)

$$\widehat{\text{Prix}} = a_0 + a_1 \cdot \text{Surface} + a_2 \cdot \text{Surface}^2 + \dots + a_{25} \cdot \text{Surface}^{25}$$

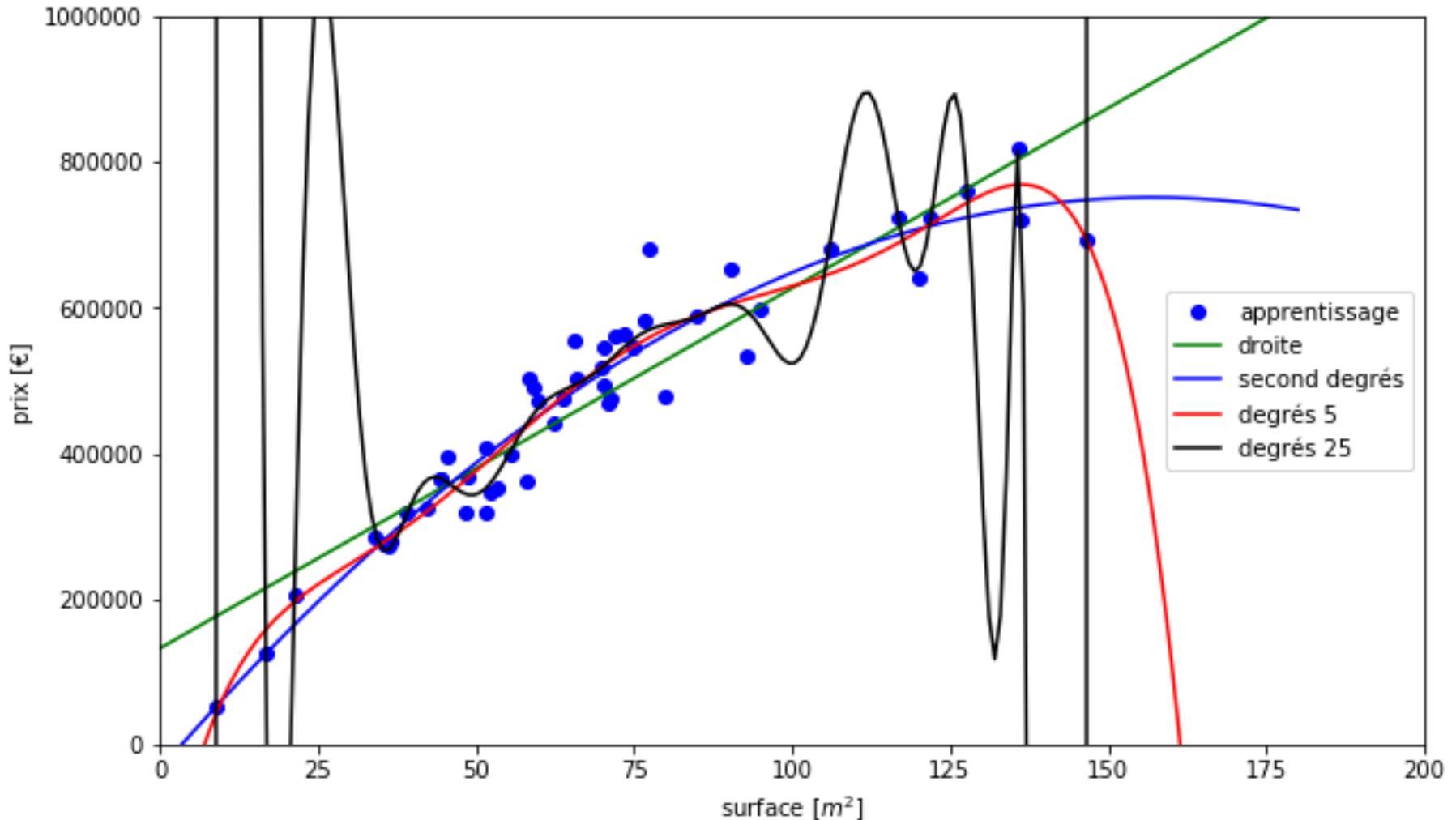
Estimateur

Degrés 25 : 26 paramètres

Nous conservons la même fonction Loss

$$L = \sum (\text{Prix} - \widehat{\text{Prix}})^2$$

Le modèle se rapproche encore plus des exemples d'apprentissage mais devient instable

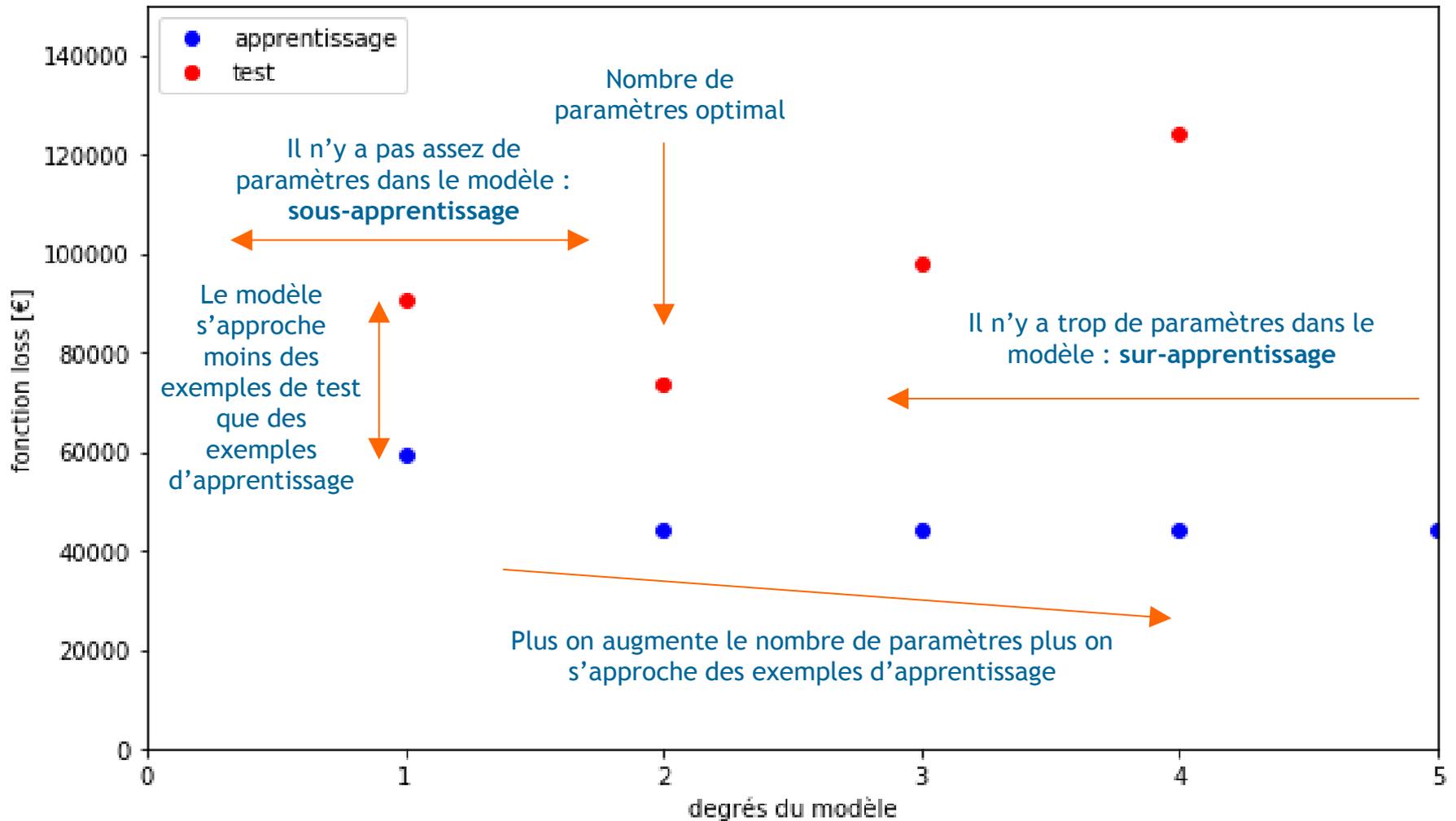




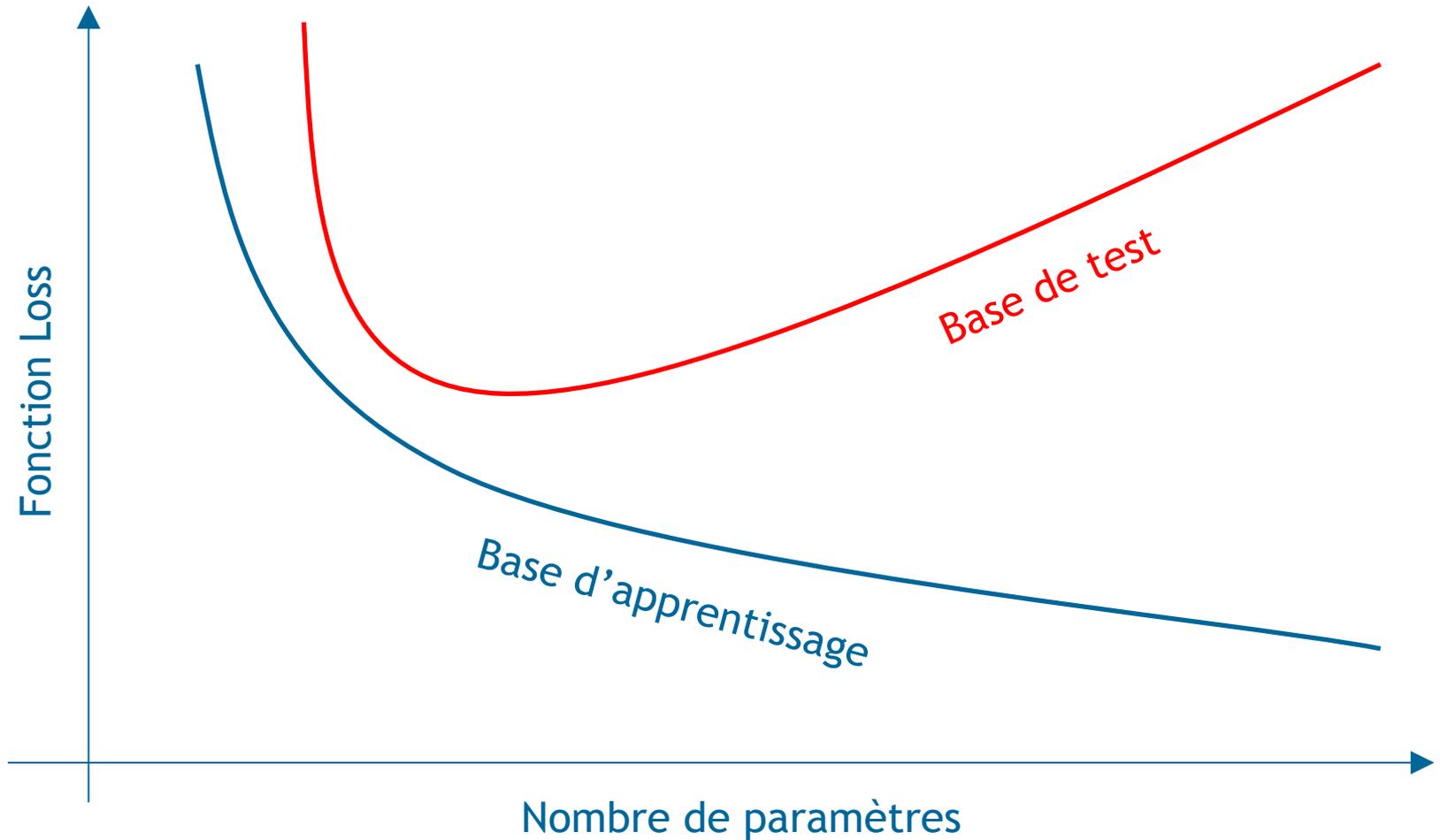
Que se passe-t-il ?

- Le modèle apprend de plus en plus parfaitement les exemples d'apprentissage
- Mais...
 - Il devient instable
 - Les paramètres prennent des valeurs positives ou négatives très grandes
- En fait le modèle apprend du « bruit »
 - Il apprend de l'information qui n'a pas de sens
- Le modèle s'éloigne des exemples de test

Il y a donc un nombre de paramètres optimal



Il s'agit d'un phénomène général appelé compromis biais-variance

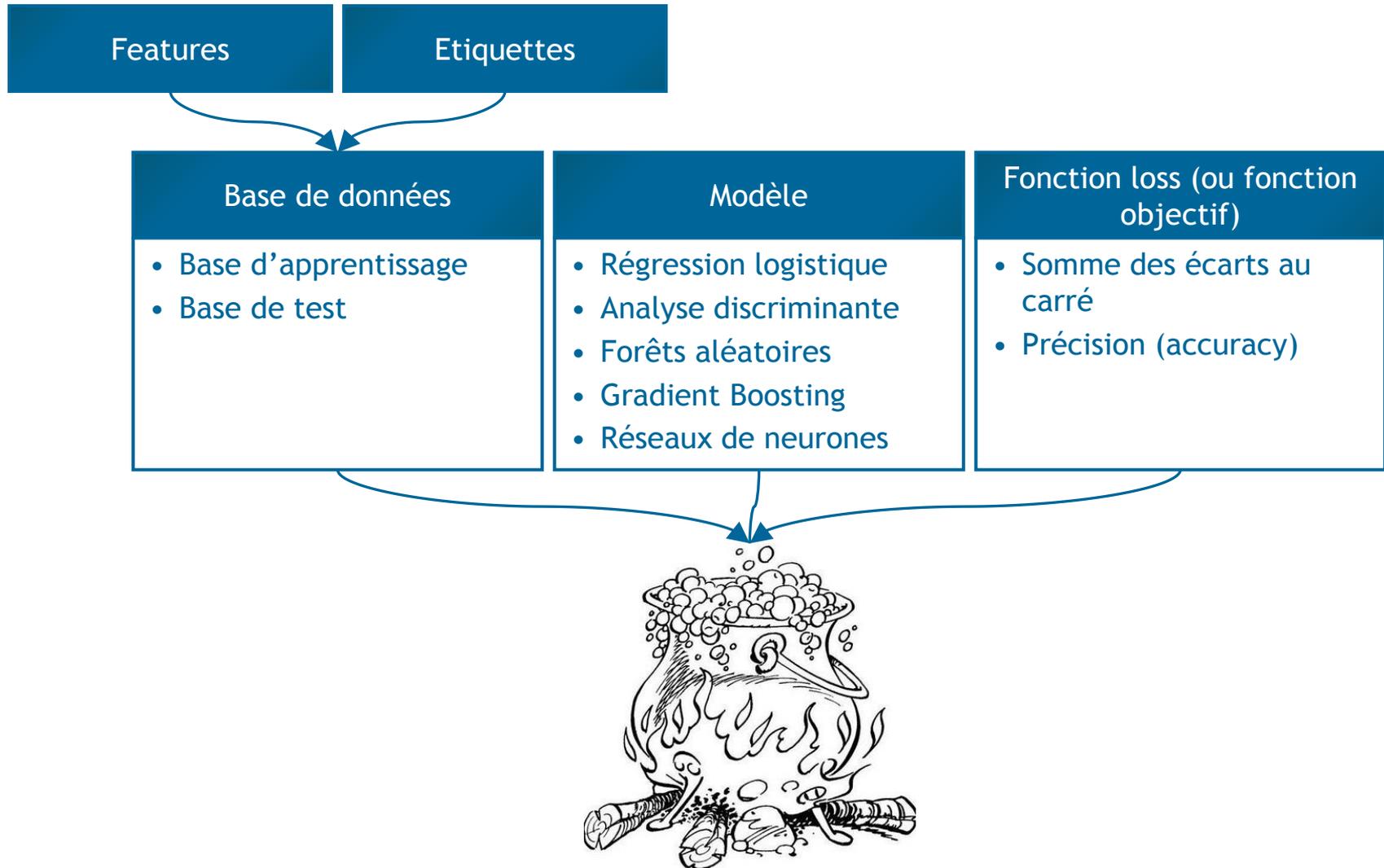




Notre liste de vocabulaire

- Machine learning ou apprentissage automatique
- Supervisé
- Non supervisé
- Exemples
- Etiquette
- Features
- Base d'apprentissage
- Base de test
- Modèle
- Paramètre
- Fonction loss ou fonction objectif
- Sous-apprentissage
- Sur-apprentissage
- Compromis biais-variance

Les 3 ingrédients d'un algorithme de machine learning





Le b.a-ba du machine learning appliqué au client : une approche basée sur la marge optimale

Deux slides de présentation

Les notions de base du machine learning

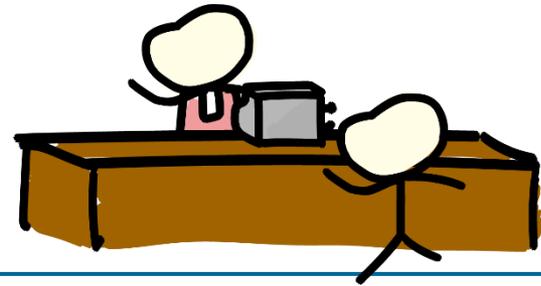
Application au crédit client : une approche basée sur la marge

Conclusion

Prenons comme exemple le risque de crédit client

Question : est ce que j'accepte de faire un crédit à mon client ?

- Paiement en 4 fois
- Paiement différé
- Paiement à la livraison
- Paiement par chèque



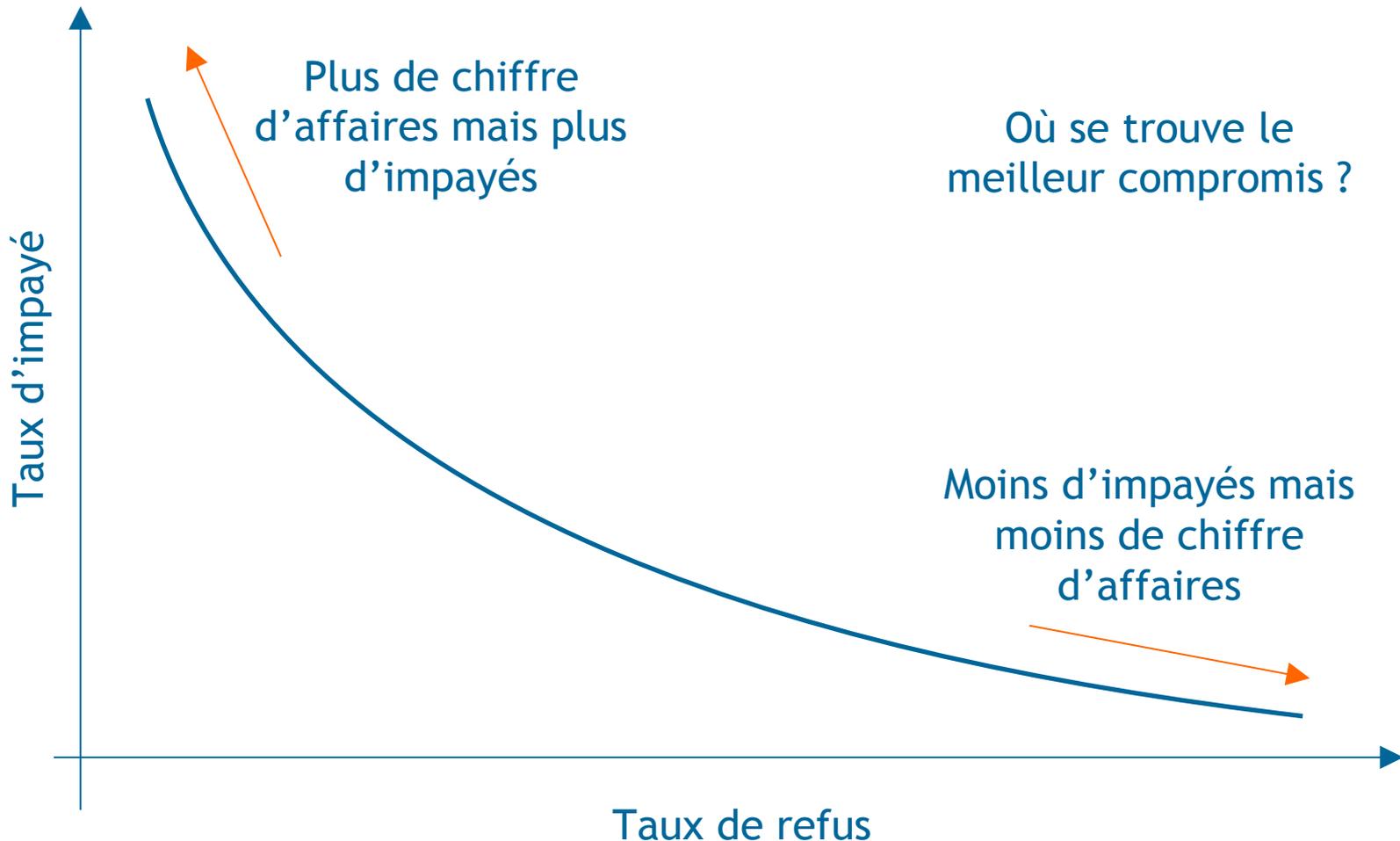
Problème de classification

- Deux classes
 - Client solvable => Crédit accordé
 - Client insolvable => Crédit refusé

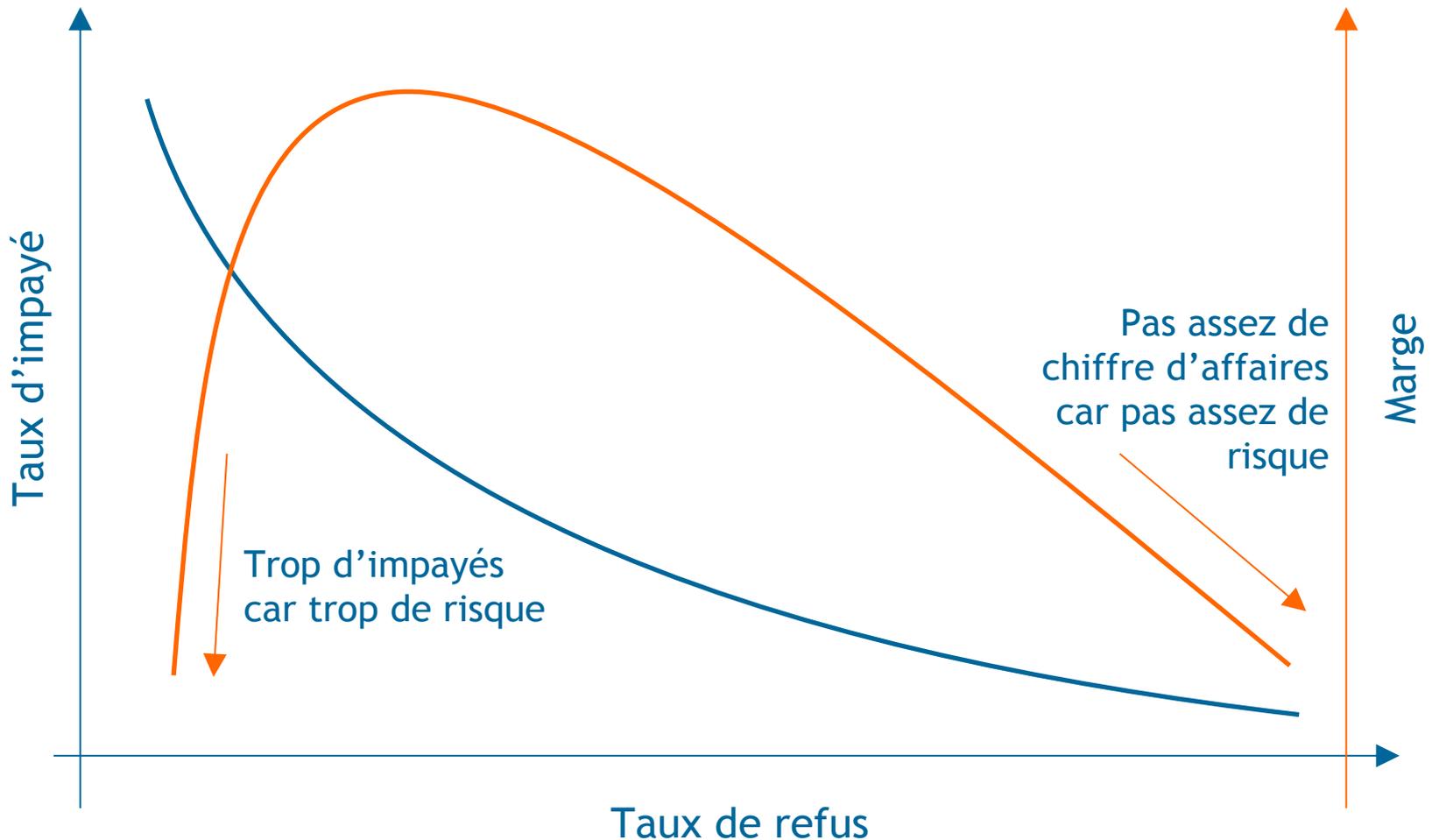
Bases de donnée

- Features
 - Contenu du panier
 - Montant
 - Magasin
 - Heure
 - Historique du client
- Etiquette
 - Client solvable / insolvable

Celui qui octroie le crédit est en face d'un dilemme entre le taux de refus et le taux d'impayé



Le meilleur compromis est celui qui optimise la marge



La marge est la fonction objectif (fonction loss)

La matrice de confusion est un outil essentiel pour analyser ce genre de problème...

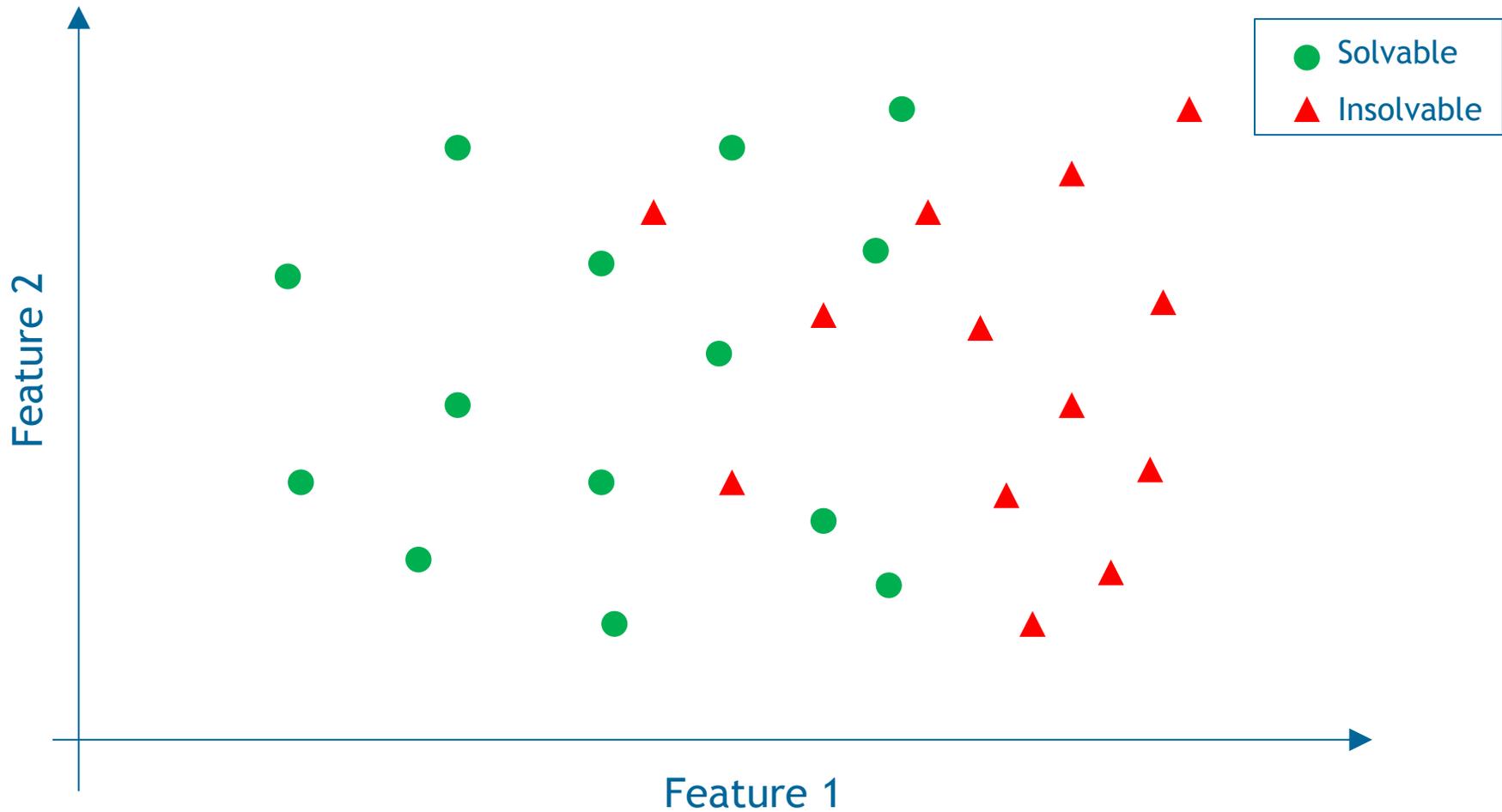
	Crédit accordé	Crédit refusé
Client solvable	n_1	n_0
Client insolvable	ε_1	ε_0

Une marge est associée à chaque situation

Marge associée	Crédit accordé	Crédit refusé
Client solvable	G_{n1}	G_{n0}
Client insolvable	$G_{\varepsilon1}$	$G_{\varepsilon0}$

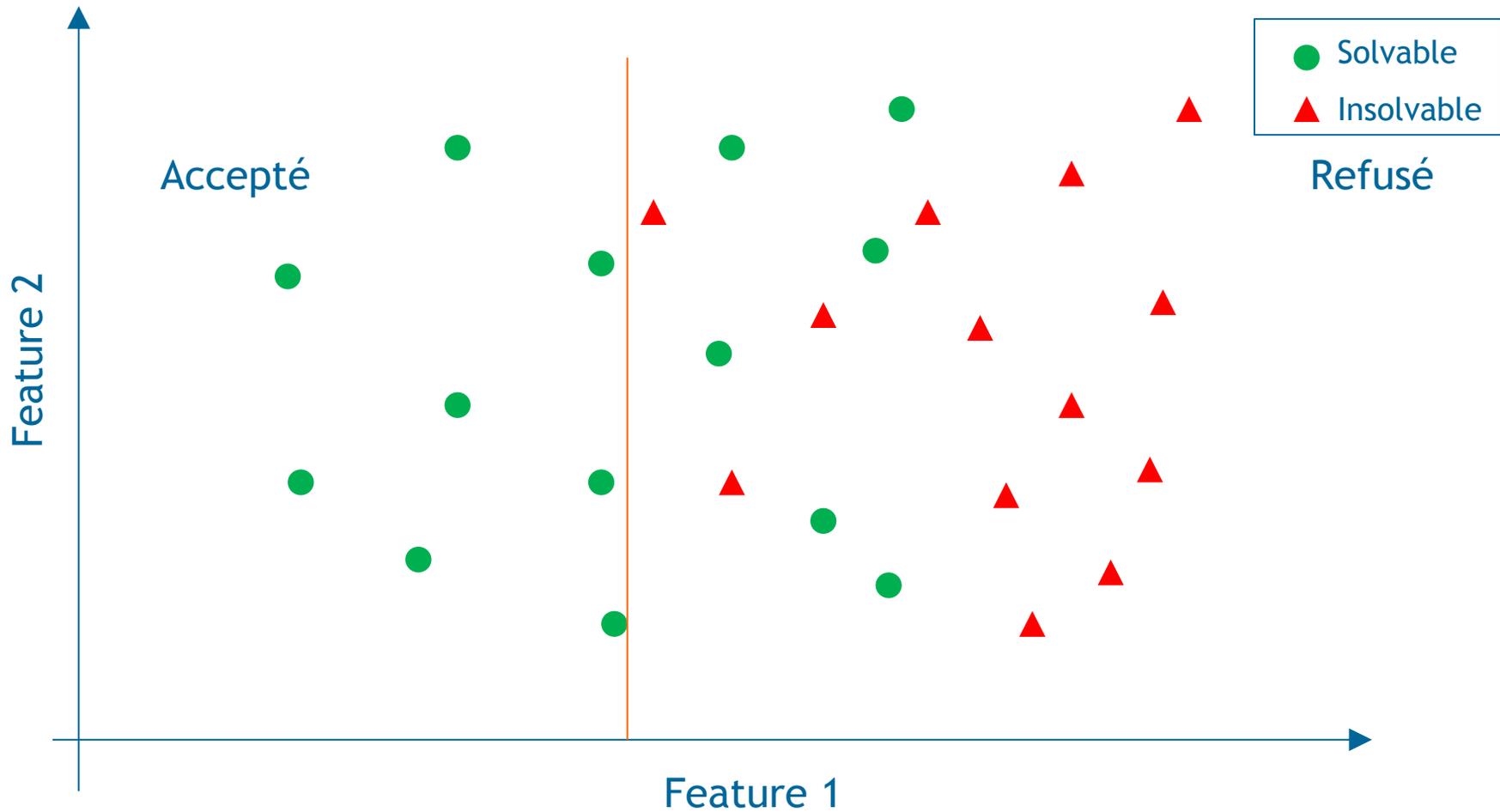
$$L = \sum G_i$$

Regardons ce qui se passe sur une base d'apprentissage avec des exemples à deux features



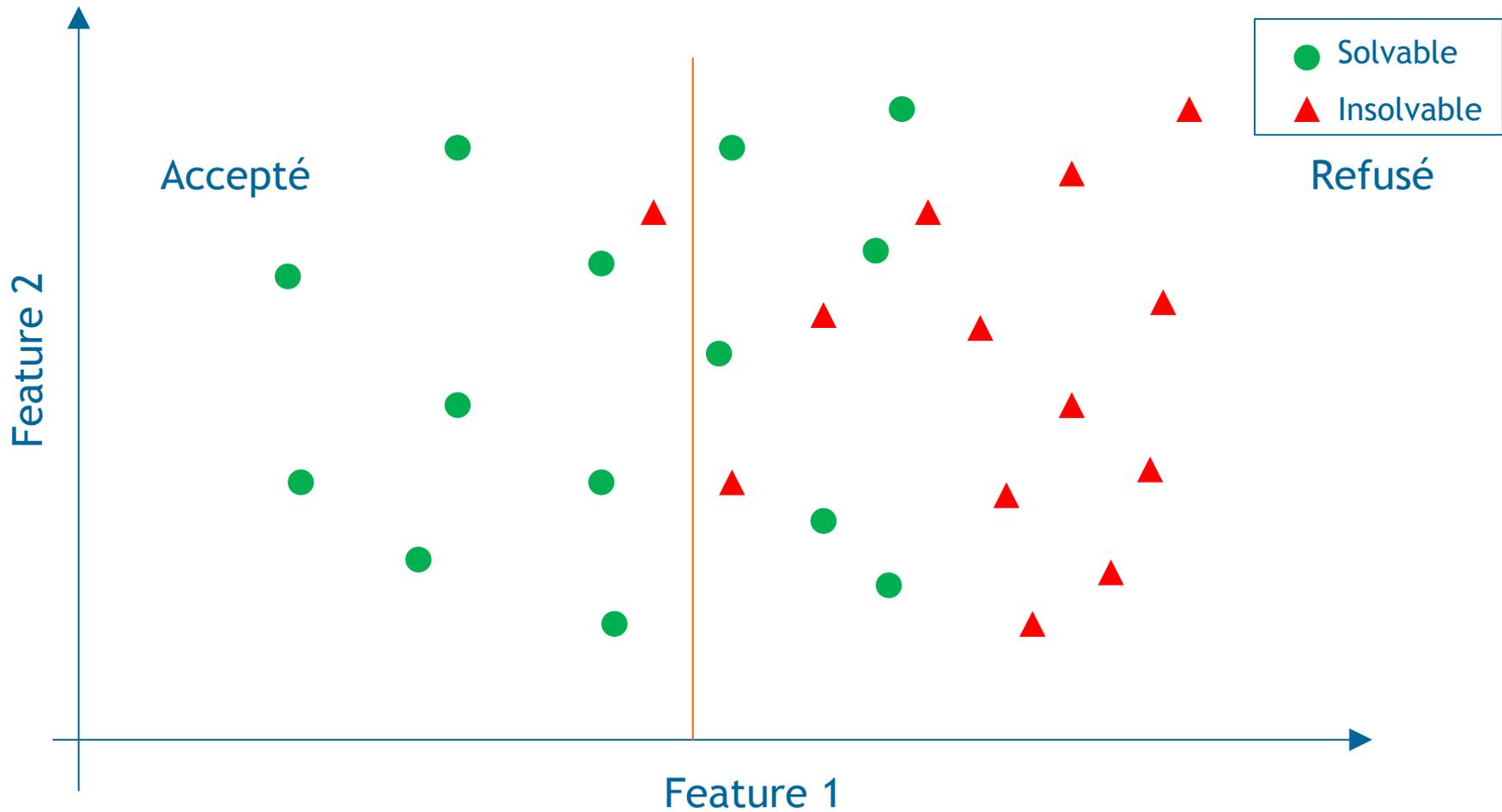
Objectif : trouver un modèle basé uniquement sur la feature 1 pour séparer les deux ensembles

La première idée est de minimiser le nombre d'exemples mal classés (fonction loss = précision = accuracy)



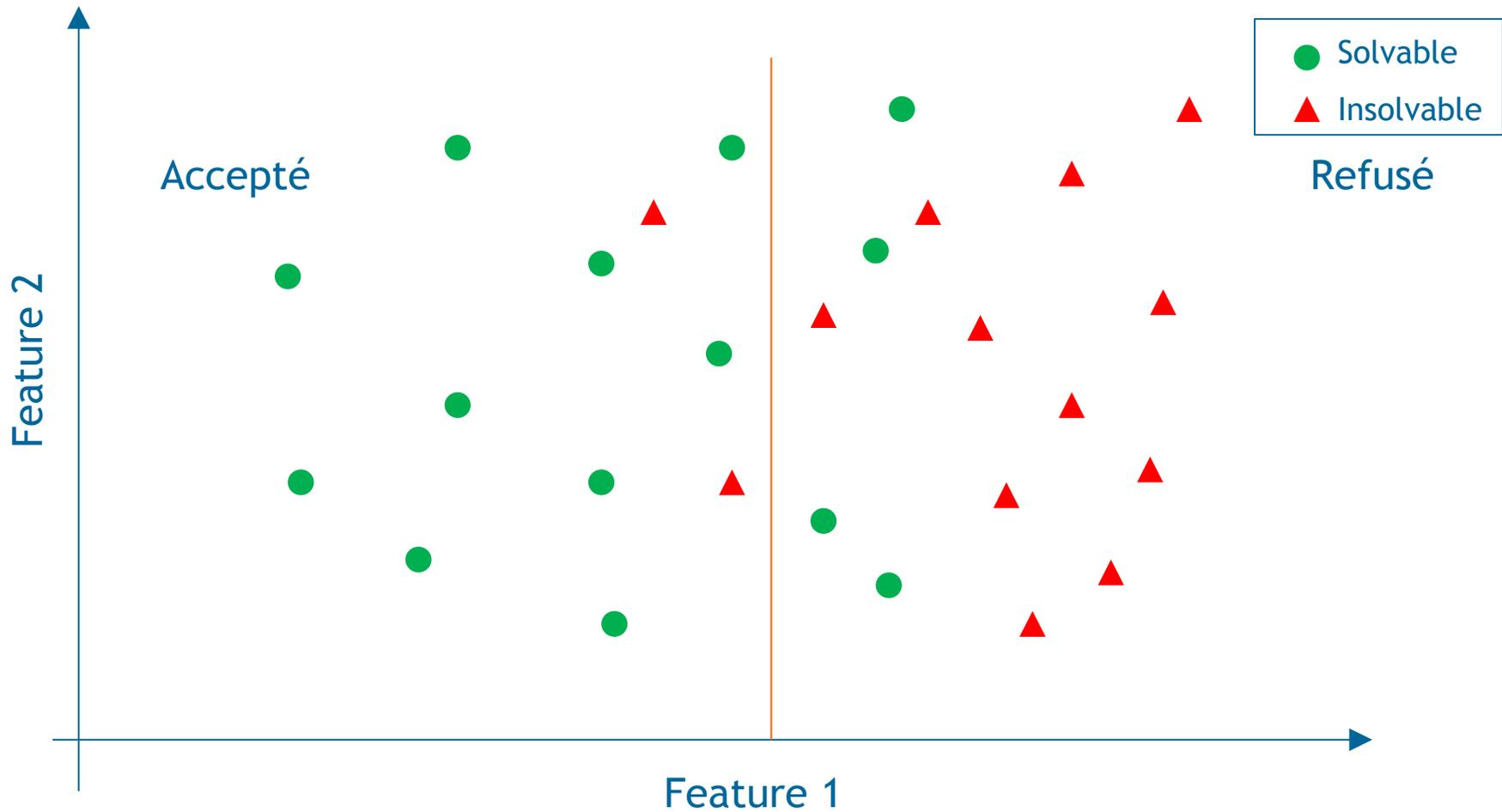
Mal classés : 6 ● + 0 ▲ = 6

La première idée est de minimiser le nombre d'exemples mal classés (fonction loss = précision = accuracy)



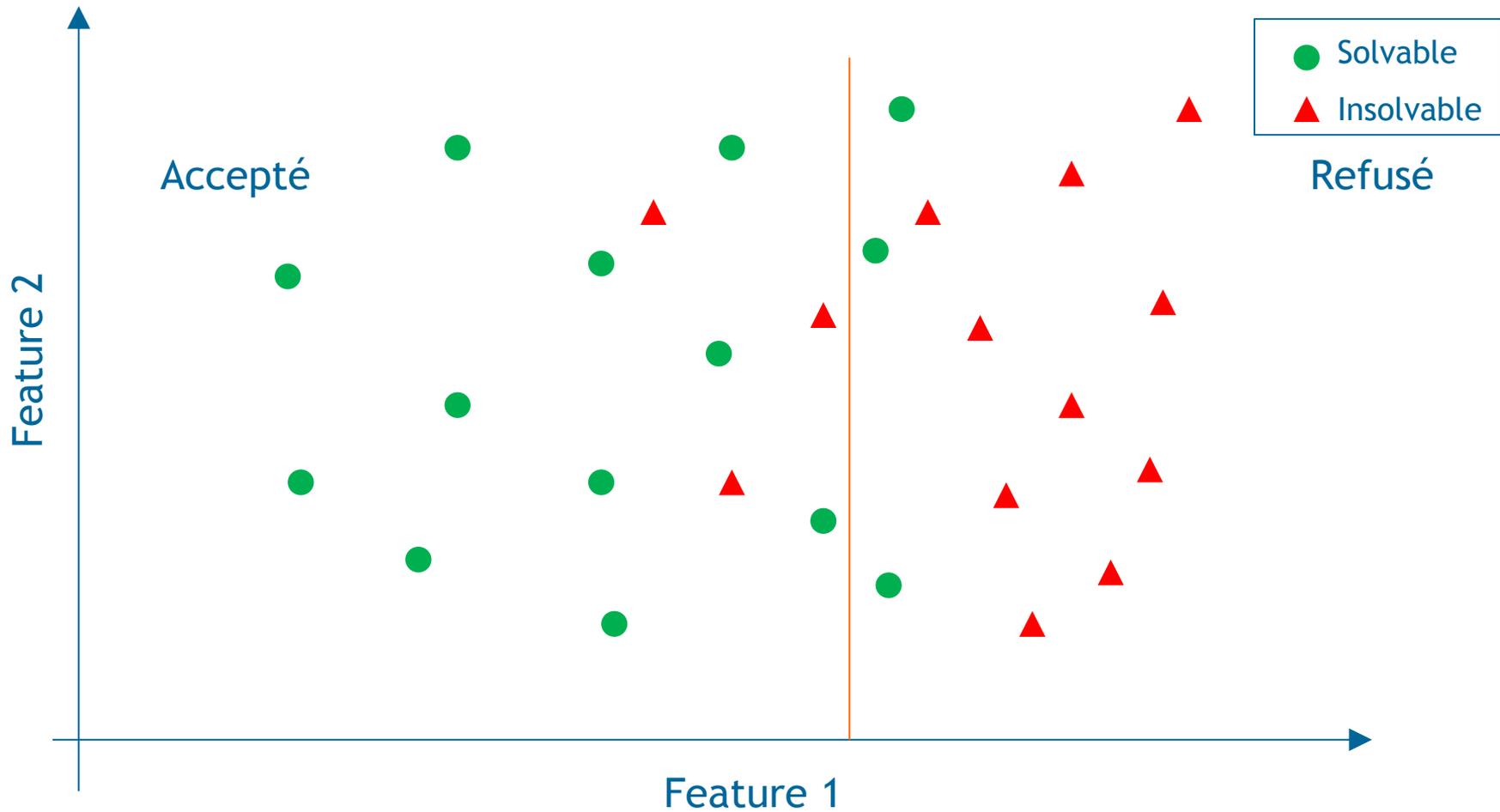
Mal classés : 6 ● + 1 ▲ = 7

La première idée est de minimiser le nombre d'exemples mal classés (fonction loss = précision = accuracy)



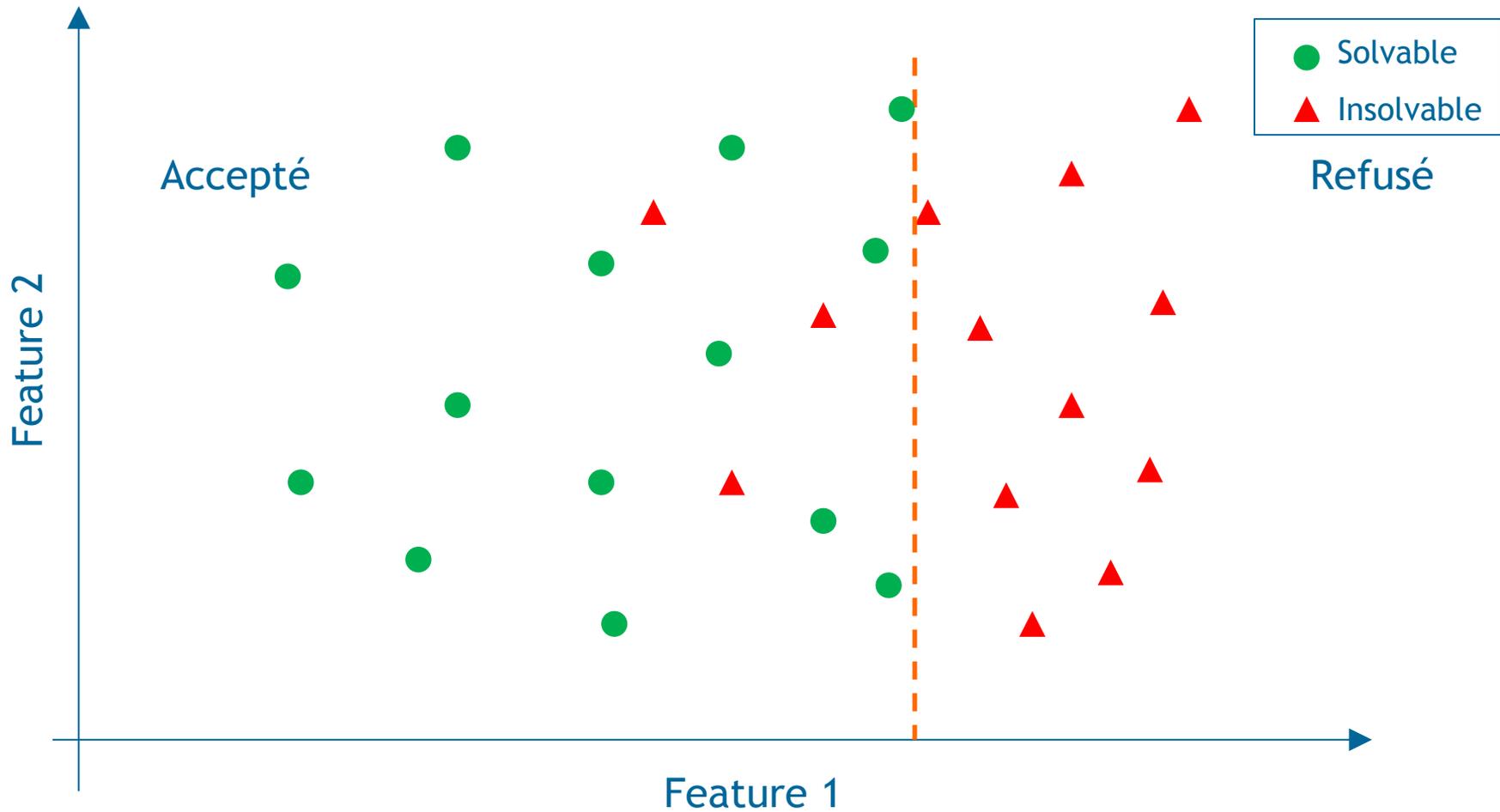
Mal classés : 4 ● + 2 ▲ = 6

La première idée est de minimiser le nombre d'exemples mal classés (fonction loss = précision = accuracy)



Mal classés : 3 ● + 3 ▲ = 6

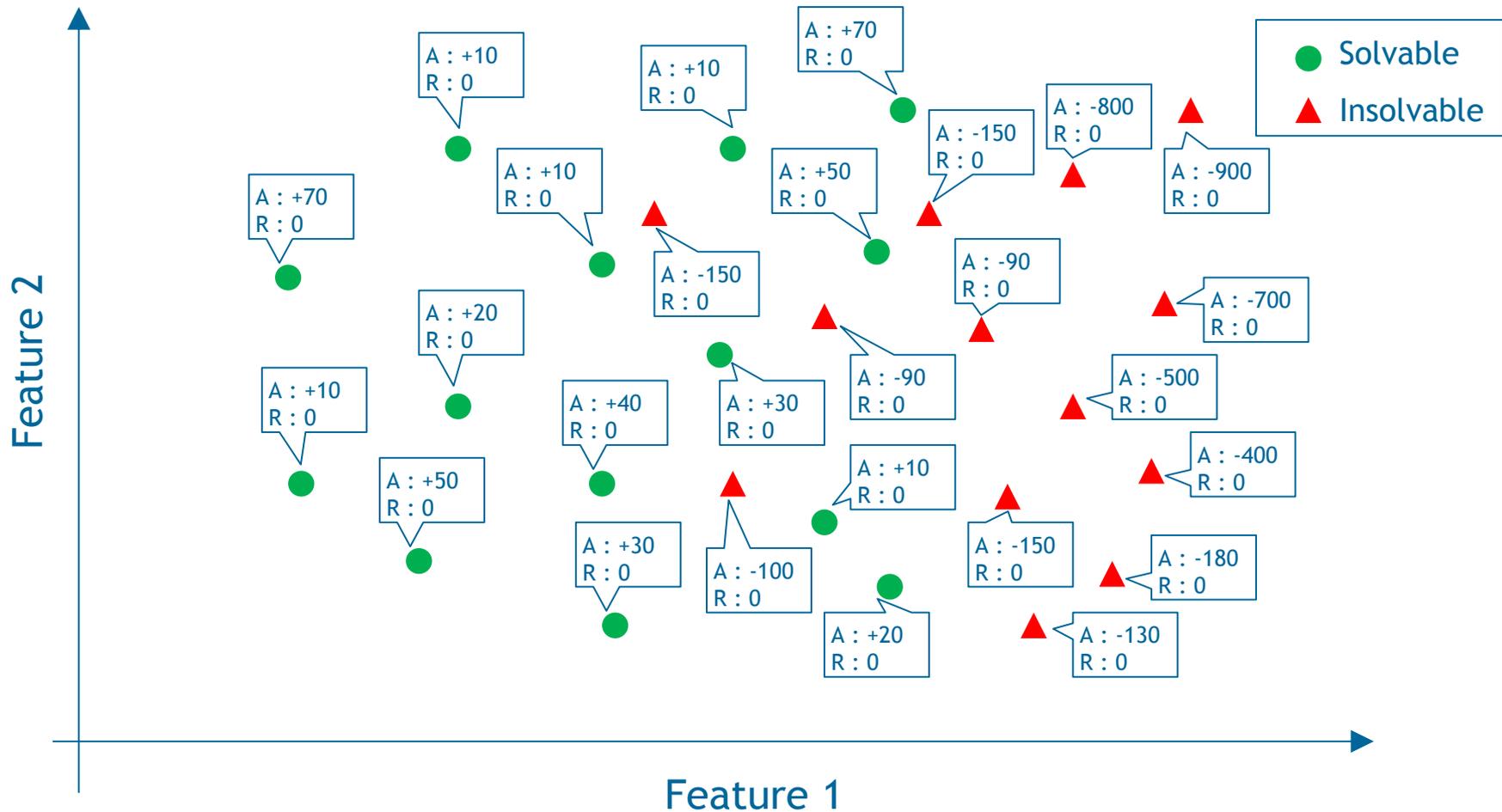
La première idée est de minimiser le nombre d'exemples mal classés (fonction loss = précision = accuracy)



Mal classés : 0 ● + 3 ▲ = 3

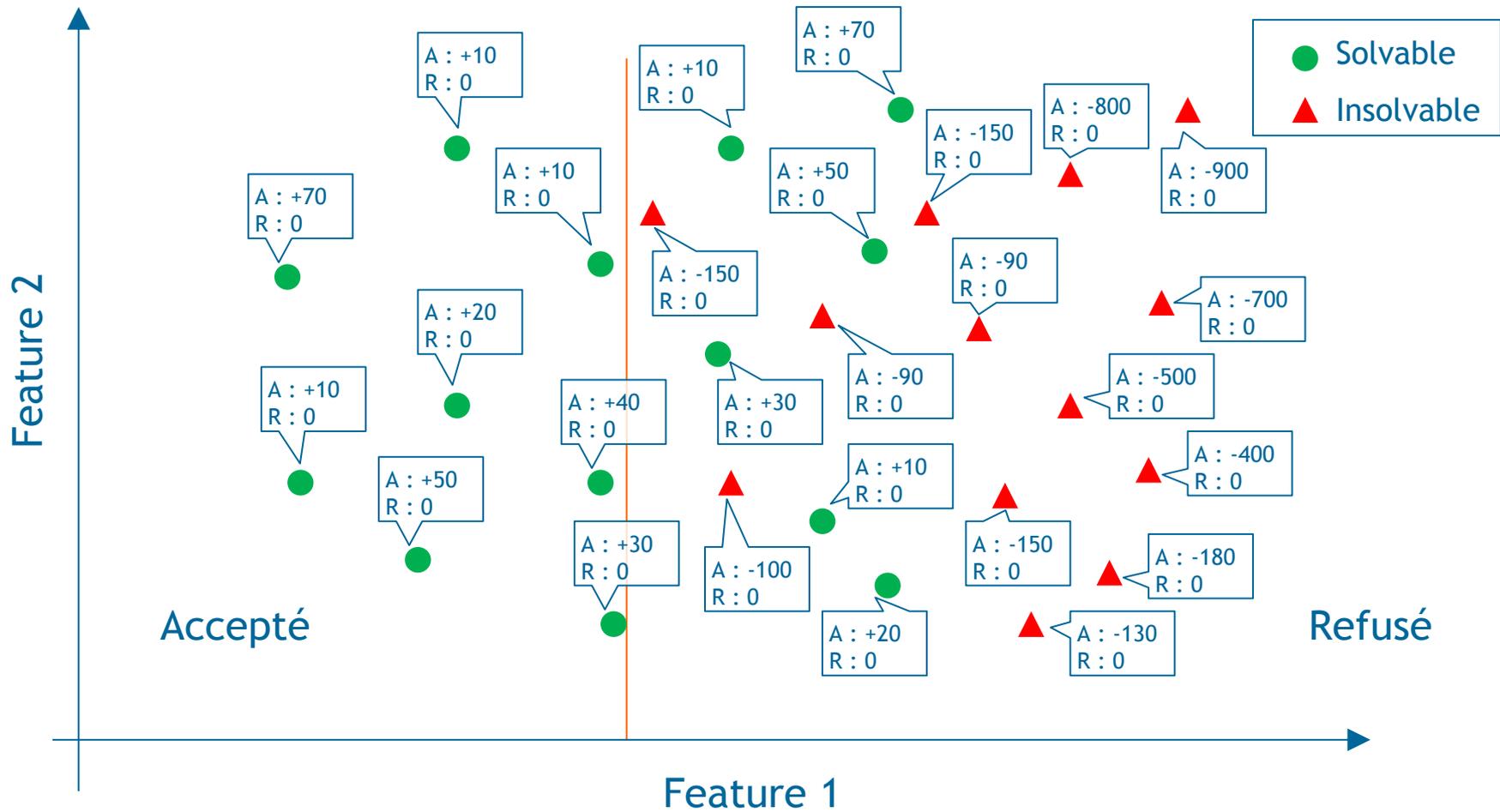
Meilleur compromis !

Prenons maintenant en compte la marge



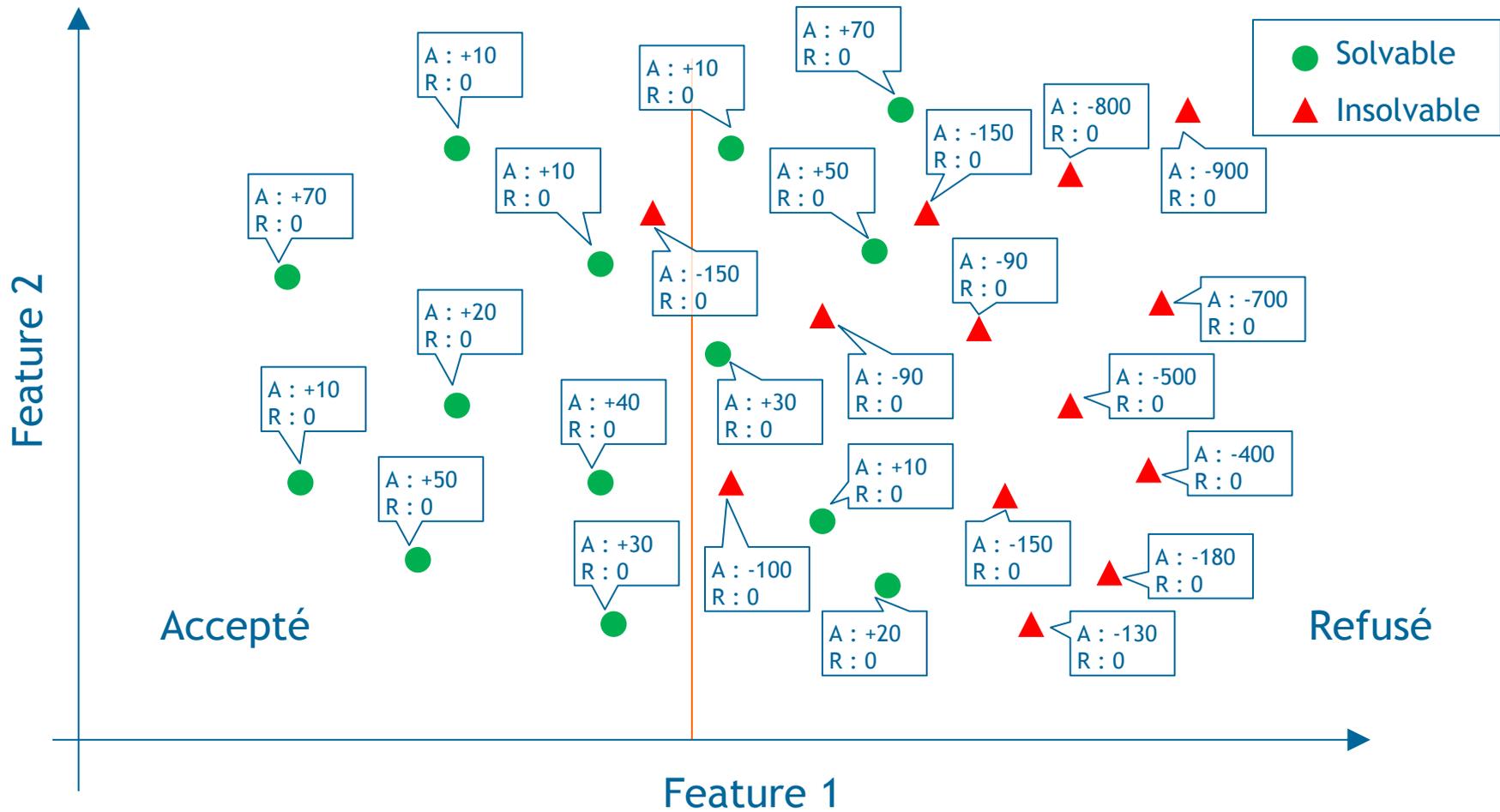
Pour chaque exemple on dispose de la marge si on l'accepte ou si on le refuse

Il est possible de chercher la marge optimale...



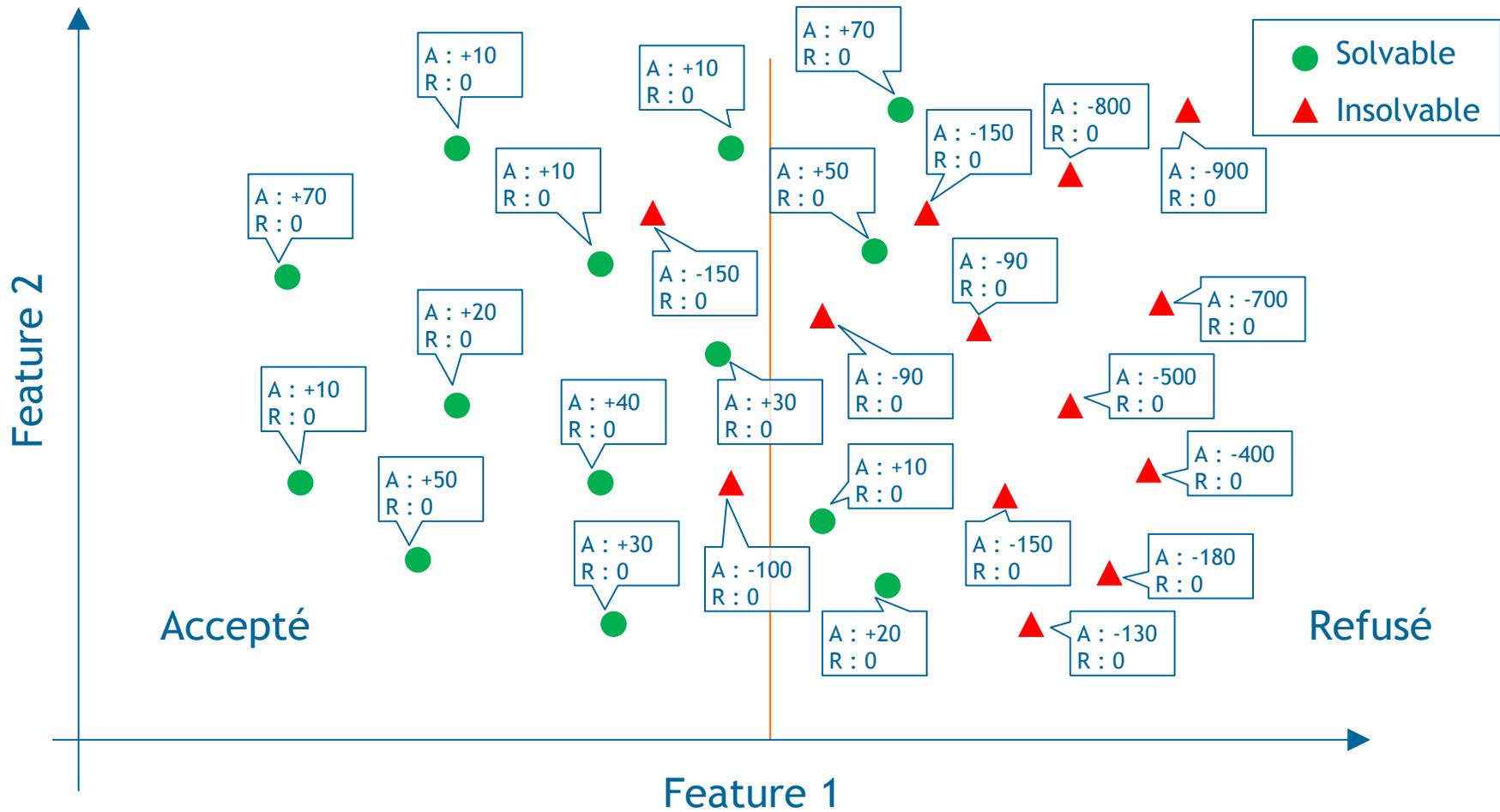
Marge : ● A : 240 + ● R : 0 + ▲ A : 0 + ▲ R : 0 = 240

Il est possible de chercher la marge optimale...



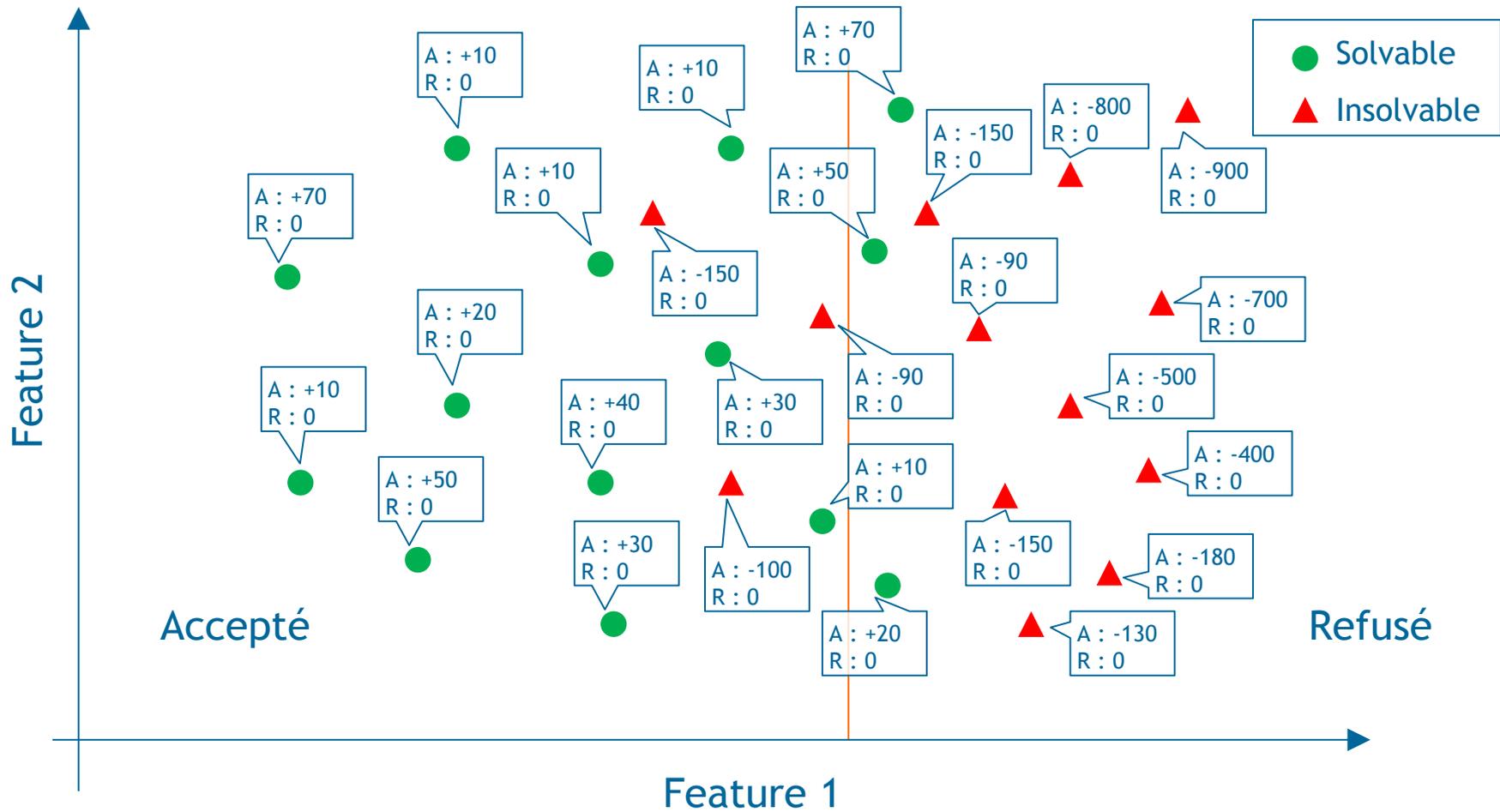
Marge : ● A : 240 + ● R : 0 + ▲ A : -150 + ▲ R : 0 = 90

Il est possible de chercher la marge optimale...



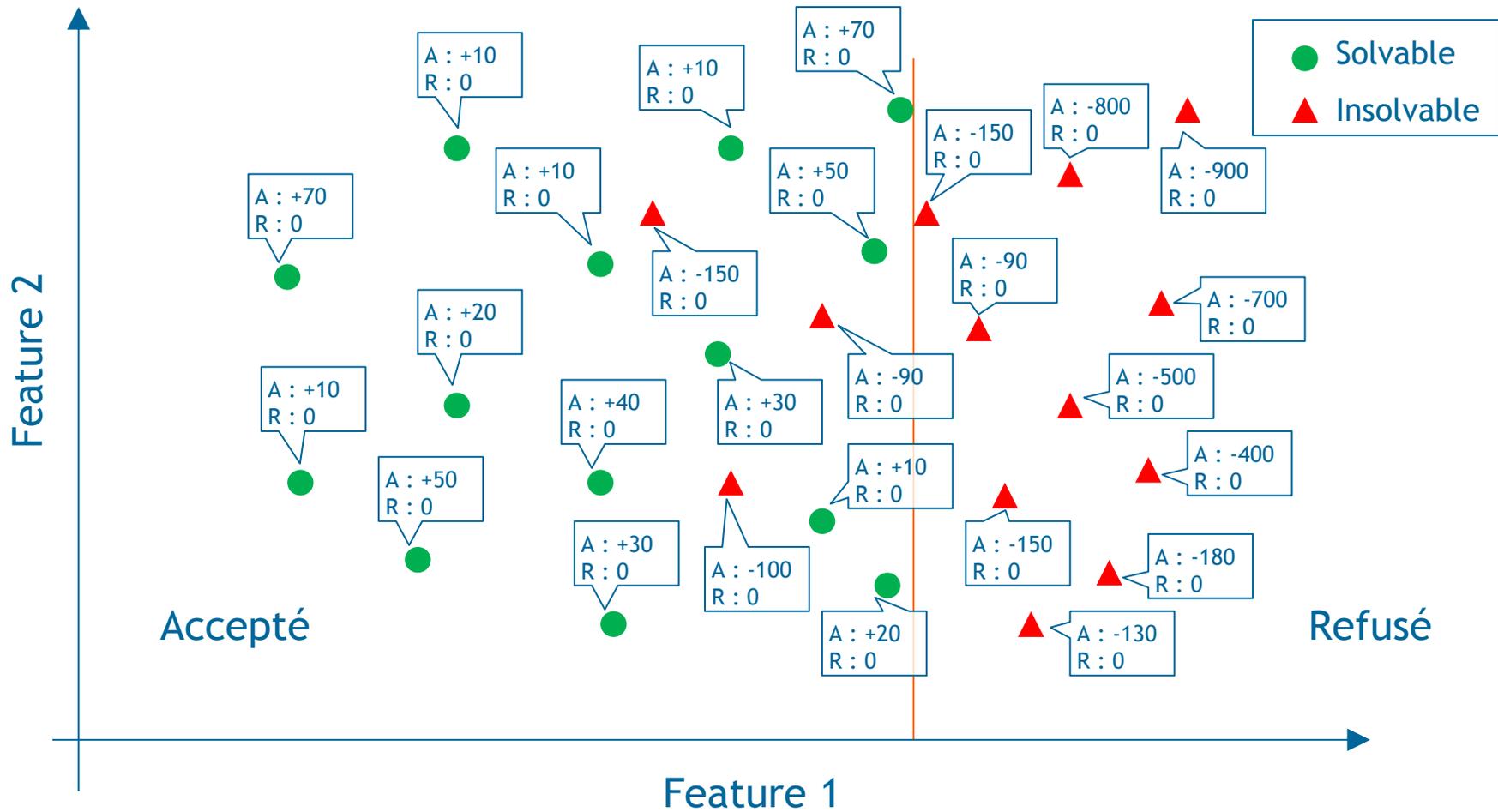
Marge : ● A : 280 + ● R : 0 + ▲ A : -250 + ▲ R : 0 = 30

Il est possible de chercher la marge optimale...



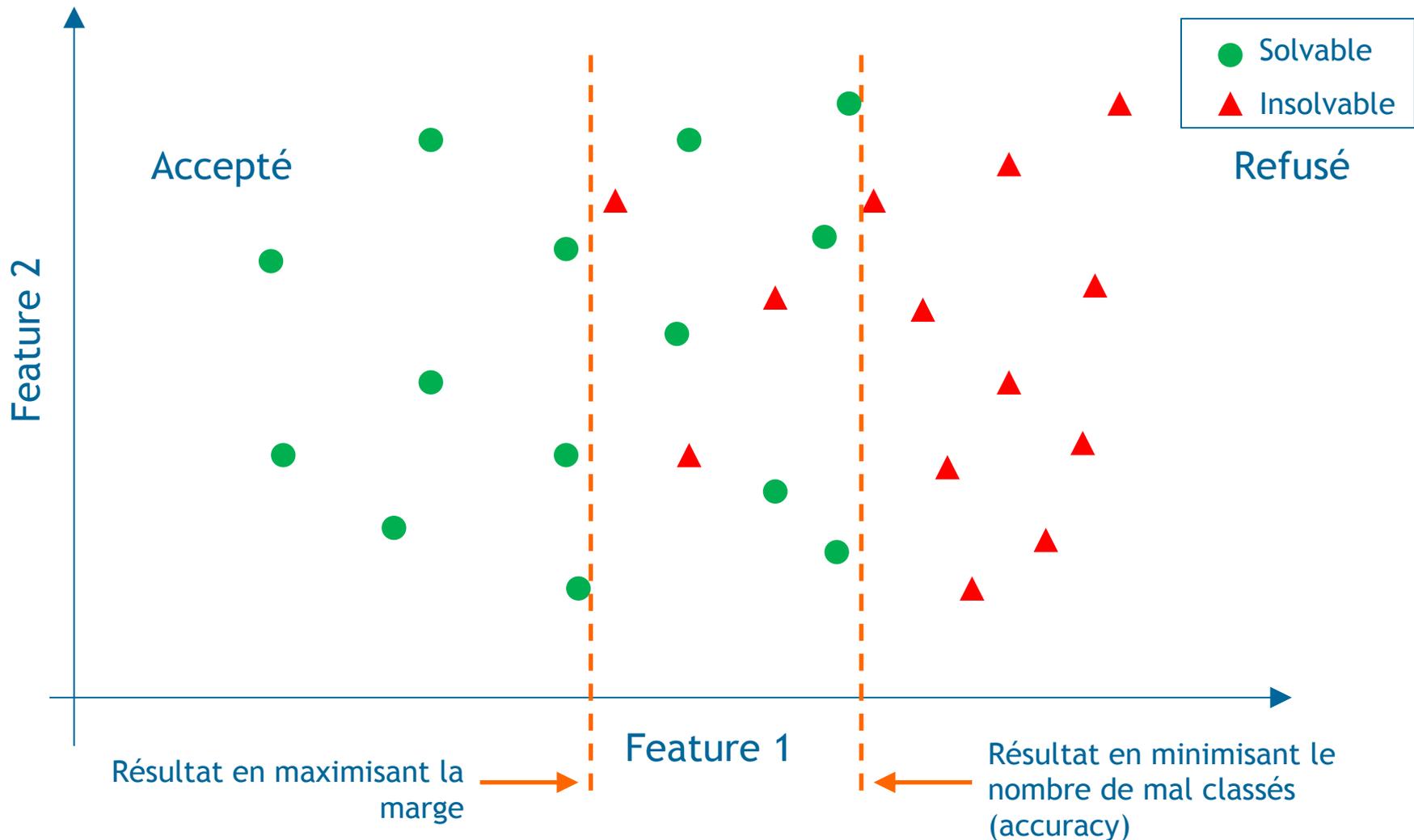
Marge : ● A : 290 + ● R : 0 + ▲ A : -340 + ▲ R : 0 = -50

Il est possible de chercher la marge optimale...

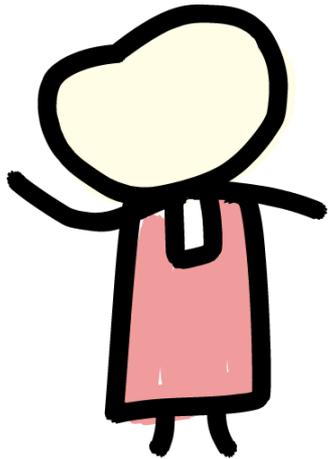


Marge : ● A : 430 + ● R : 0 + ▲ A : -340 + ▲ R : 0 = 90

Avec la même base et le même modèle, une fonction loss différente conduit à un résultat différent



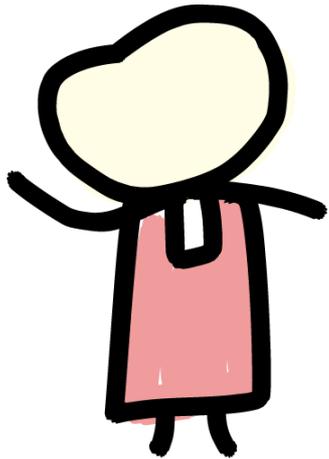
Construisons le modèle de marge d'un commerçant



	Crédit accordé	Crédit refusé
Solvable	r.m	
Insolvable		

- m : montant de la transaction
- r : taux de marge

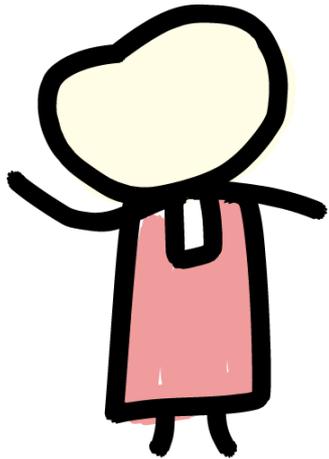
Construisons le modèle de marge d'un commerçant



	Crédit accordé	Crédit refusé
Solvable	$r.m$	
Insolvable		0

- m : montant de la transaction
- r : taux de marge

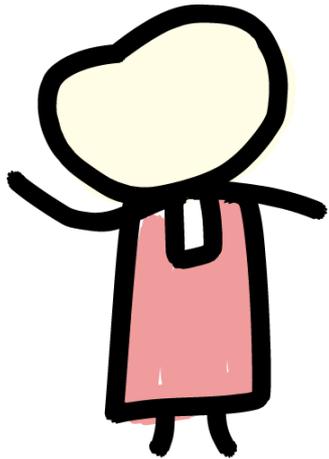
Construisons le modèle de marge d'un commerçant



	Crédit accordé	Crédit refusé
Solvable	$r.m$	
Insolvable	$(r-c).m$	0

- m : montant de la transaction
- r : taux de marge
- c : coût des restant-dus et coûts des opérations de recouvrement

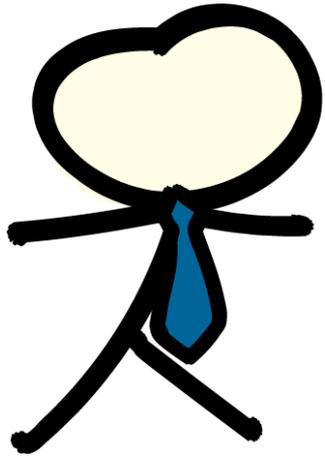
Construisons le modèle de marge d'un commerçant



	Crédit accordé	Crédit refusé
Solvable	$r.m$	$\rho.r.m - p$
Insolvable	$(r-c).m$	0

- m : montant de la transaction
- r : taux de marge
- c : coût des restant-dus et coûts des opérations de recouvrement
- ρ : taux de poursuite sans crédit client
- p : coût associé à un client mécontent d'avoir la facilité de paiement refusée

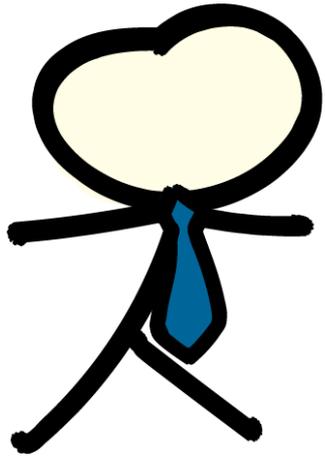
Construisons le modèle de marge d'une banque qui octroie le crédit



	Crédit accordé	Crédit refusé
Solvable	$\pi \cdot m$	
Insolvable		

- m : montant de la transaction
- π : commission de garantie

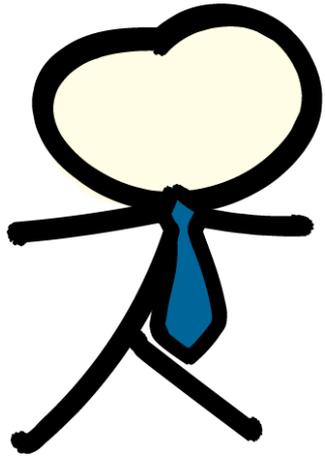
Construisons le modèle de marge d'une banque qui octroie le crédit



	Crédit accordé	Crédit refusé
Solvable	$\pi \cdot m$	
Insolvable		0

- m : montant de la transaction
- π : commission de garantie

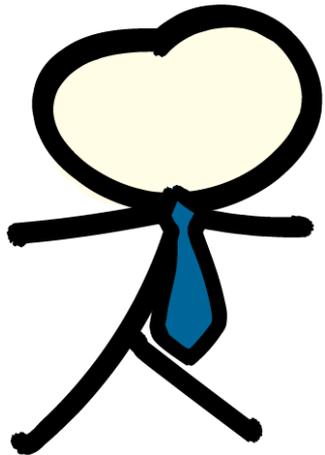
Construisons le modèle de marge d'une banque qui octroie le crédit



	Crédit accordé	Crédit refusé
Solvable	$\pi \cdot m$	0
Insolvable		0

- m : montant de la transaction
- π : commission de garantie

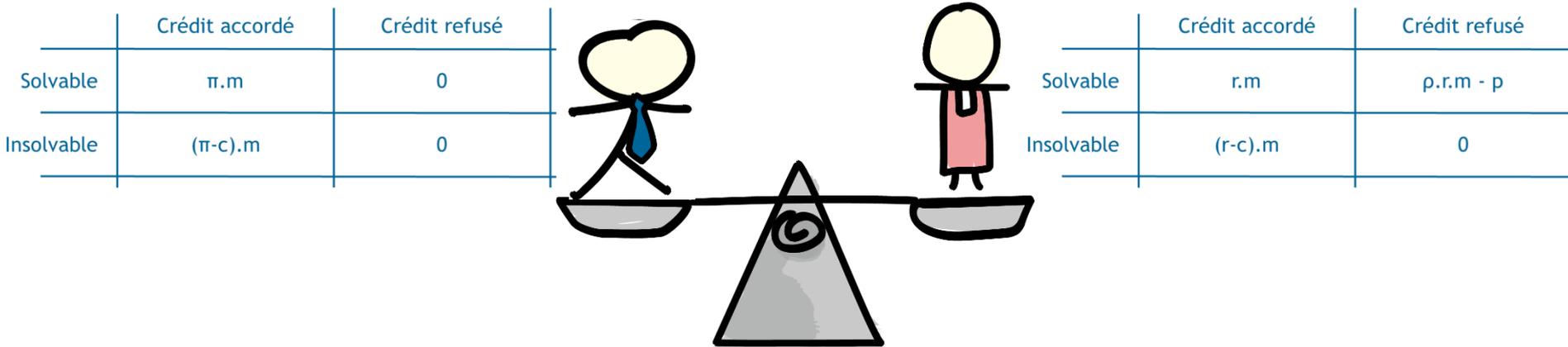
Construisons le modèle de marge d'une banque qui octroie le crédit



	Crédit accordé	Crédit refusé
Solvable	$\pi.m$	0
Insolvable	$(\pi-c).m$	0

- m : montant de la transaction
- π : commission de garantie
- c : coût des restant-dus et coûts des opérations de recouvrement

Les intérêts du commerçant et de la banque ne sont pas alignés



- Avec la même base de données et les mêmes modèles, la banque et le commerçant vont faire des choix différents
- La banque va refuser des crédits que le commerçant aurait intérêt à accepter

Prenons un exemple très simple pour sentir ce qui se passe

- Quand le crédit est refusé le client abandonne systématiquement la vente
 - $\rho = 0$
- En cas d'impayé, il n'y a aucun recouvrement. Le restant dû est égal au montant de la transaction
 - $c = 100\%$
- Un client souhaite un crédit pour un panier de 100 €
 - $m = 100 \text{ €}$
- Le refus de crédit n'a aucun impact au-delà de la transaction
 - $p = 0 \text{ €}$
- Une banque propose de couvrir le risque avec une commission de 3,5%
 - $\pi = 3,5\%$

1^{er} cas :

Marge du panier < Commission de garantie < Risque d'impayé

		Marge du panier	
		2%	15%
Risque d'impayé	3%		
	10%	Cas n° 1	

Le commerçant choisit de couvrir son risque

Le commerçant prend son risque

		Marge du commerçant	
		Crédit	
		Accordé	Refusé
Solvable	90%	2,0 €	0 €
Insolvable	10%	-98,0 €	0 €
Total		-8,0 €	0 €

		Marge de la banque	
		Crédit	
		Accordé	Refusé
90%		3,5 €	0 €
10%		-96,5 €	0 €
		-6,5 €	0 €

		Marge restante au commerçant	
		Crédit	
		Accordé	Refusé
90%		-1,5 €	0 €
10%		-1,5 €	0 €
		-1,5 €	0 €



2^{ème} cas :

Marge du panier < Risque d'impayé < Commission de garantie

		Marge du panier	
		2%	15%
Risque d'impayé	3%	Cas n° 2	
	10%		

Le commerçant choisit de couvrir son risque

Le commerçant prend son risque



		Marge du commerçant	
		Crédit	
		Accordé	Refusé
Solvable	97%	2,0 €	0 €
Insolvable	3%	-98,0 €	0 €
Total		-1 €	0 €



		Marge de la banque	
		Crédit	
		Accordé	Refusé
Solvable	97%	3,5 €	0 €
Insolvable	3%	-96,5 €	0 €
Total		0,5 €	0 €



		Marge restante au commerçant	
		Crédit	
		Accordé	Refusé
Solvable	97%	-1,5 €	0 €
Insolvable	3%	-1,5 €	0 €
Total		-1,5 €	0 €



3^{ème} cas :

Risque d'impayé < Commission de garantie < Marge du panier

		Marge du panier	
		2%	15%
Risque d'impayé	3%		Cas n° 3
	10%		

Le commerçant choisit de couvrir son risque

Le commerçant prend son risque



		Marge du commerçant	
		Crédit	
		Accordé	Refusé
Solvable	97%	15,0 €	0 €
Insolvable	3%	-85,0 €	0 €
Total		12,0 €	0 €



		Marge de la banque	
		Crédit	
		Accordé	Refusé
Solvable	97%	3,5 €	0 €
Insolvable	3%	-96,5 €	0 €
Total		0,5 €	0 €



		Marge restante au commerçant	
		Crédit	
		Accordé	Refusé
Solvable	97%	11,5 €	0 €
Insolvable	3%	11,5 €	0 €
Total		11,5 €	0 €



4^{ème} cas :

Commission de garantie < Risque d'impayé < Marge du panier

		Marge du panier	
		2%	15%
Risque d'impayé	3%		
	10%		Cas n° 4

Le commerçant choisit de couvrir son risque

Le commerçant prend son risque

		Marge du commerçant			Marge de la banque			Marge restante au commerçant	
		Crédit			Crédit			Crédit	
		Accordé	Refusé		Accordé	Refusé		Accordé	Refusé
Solvable	90%	15,0 €	0 €	90%	3,5 €	0 €	90%	11,5 €	0 €
Insolvable	10%	-85,0 €	0 €	10%	-96,5 €	0 €	10%	11,5 €	0 €
Total		5,0 €	0 €		-6,5 €	0 €		11,5 €	0 €

Note: Green arrows point to the 5,0 € and 11,5 € values in the Total row. An orange bracket spans the bank and remaining merchant margin tables.

A information égale entre la banque et le commerçant, ce dernier n'a jamais intérêt à couvrir son risque client

Cas 1 : la commission est trop chère par rapport à la marge

Marge du panier



Commission de garantie



Cas 2 : la commission n'est pas assez chère ! La banque ne prendra pas assez de risque

Commission de garantie



Marge du panier



Les clients dont le risque est compris entre la commission de garantie et la marge du panier vont être refusés

La situation est plus complexe car il faut tenir compte du taux de poursuite et de la capacité à recouvrer les impayés



Le b.a-ba du machine learning appliqué au client : une approche basée sur la marge optimale

Deux slides de présentation

Les notions de base du machine learning

Application au crédit client : une approche basée sur la marge

Conclusion

Quelques messages à retenir...

1

Les 3 ingrédients du Machine Learning sont les bases de données, le modèle et la fonction objectif

2

La marge est une bonne fonction objectif

3

Un bon modèle de marge prend en compte les opérations de bout en bout : de la marge du panier au taux de recouvrement

4

Deux acteurs disposant d'une même base de données et d'un même modèle de machine learning feront des choix différents s'ils utilisent un modèle de marge différent

5

Les intérêts d'un commerçant et d'une banque ne sont pas alignés