Retour d'expérience sur l'enseignement de l'apprentissage automatique à destination d'ingénieurs généralistes

L'exemple de Mines Nancy

21 septembre 2023
Arts et Métiers-Campus de Paris
ENSAM

Frédéric SUR

Laboratoire lorrain de recherche en informatique et ses applications – LORIA UMR 7503
CNRS, Inria, Université de Lorraine

Journée Intelligence artificielle et génie civil

« Introduction à l'apprentissage automatique »

- Cours de **tronc commun scientifique** « ingénieur civil » de **Mines Nancy**
  Départements de Mines Nancy: *Informatique, Génie Industriel et Mathématiques Appliquées, Energie, Sciences et Ingénierie des Matériaux, Géosciences et Génie Civil*

- **6 séances** (1h CM + 2hTD) en 2ème année (Bac+4), novembre à janvier + **un examen**

- Depuis novembre 2020… (à distance en 2020)

- Plutôt que « IA »: **apprentissage automatique** (*machine learning*)

Motivations

science & médecine Le Monde

L'héritage immunitaire de la peste noire

L'hécatombe du XIVe siècle a sélectionné un gène protecteur contre « Yersinia pestis », qui aujourd'hui accroît le risque de pathologies auto-immunes

PAGE 2

Entretien
Le scandale des « hectares de peau »

De 1951 à 1974, des détenus afro-américains ont servi de cobayes dans des expériences de dermatologie. Allen Hornblum a dénoncé ces pratiques

PAGE 8

L'intelligence artificielle, nouveau moteur de la recherche

Cette technologie d'apprentissage par les données envahit tous les champs de la science. Parce que l'IA permet d'améliorer les calculs et d'affiner les instruments, elle a fini par convaincre même les plus dubitatifs des chercheurs

DAVID LAROUSSERIE

À NOS LECTEURS

- **PNRIA 2018** (plan national de recherche en intelligence artificielle),
- **rapport « Cédric Villani »**,
- **offres de stages et d'emploi**,
- **IA partout**

(assistants vocaux, traduction automatique, algo de recommandation…)

Le Monde, 26 octobre 2022:

« Cette technologie d'apprentissage par les données envahit tous les champs de la science. Parce que l'IA permet d'améliorer les calculs et d'affiner les résultats, elle a fini par convaincre même les plus dubitatifs des chercheurs »

Exemple: un problème « physique »

Question:

21 septembre 2023
Arts et Métiers-Campus de Paris
ENSAM

- Ma voiture roule à 72 km/h.
- Quelle est la distance d'arrêt ?

Solution 1 : méthode expérimentale

→ on fait l'expérience…

21 septembre 2023
Arts et Métiers-Campus de Paris
ENSAM

Question 2 :

- Ma voiture roule à 105,77 km/h.
- Quelle est la distance d'arrêt ?
- etc.

Inconvénients : expérience possible ? coût ?

Solution 2 : modélisation physique

➡ approche « basée modèle » :

21 sep
Arts et
ENSAM

$$D = v_0 t_r + \frac{v_0^2}{2a}$$

où *a* décélération, *v0* vitesse initiale, *tr* temps de réaction.

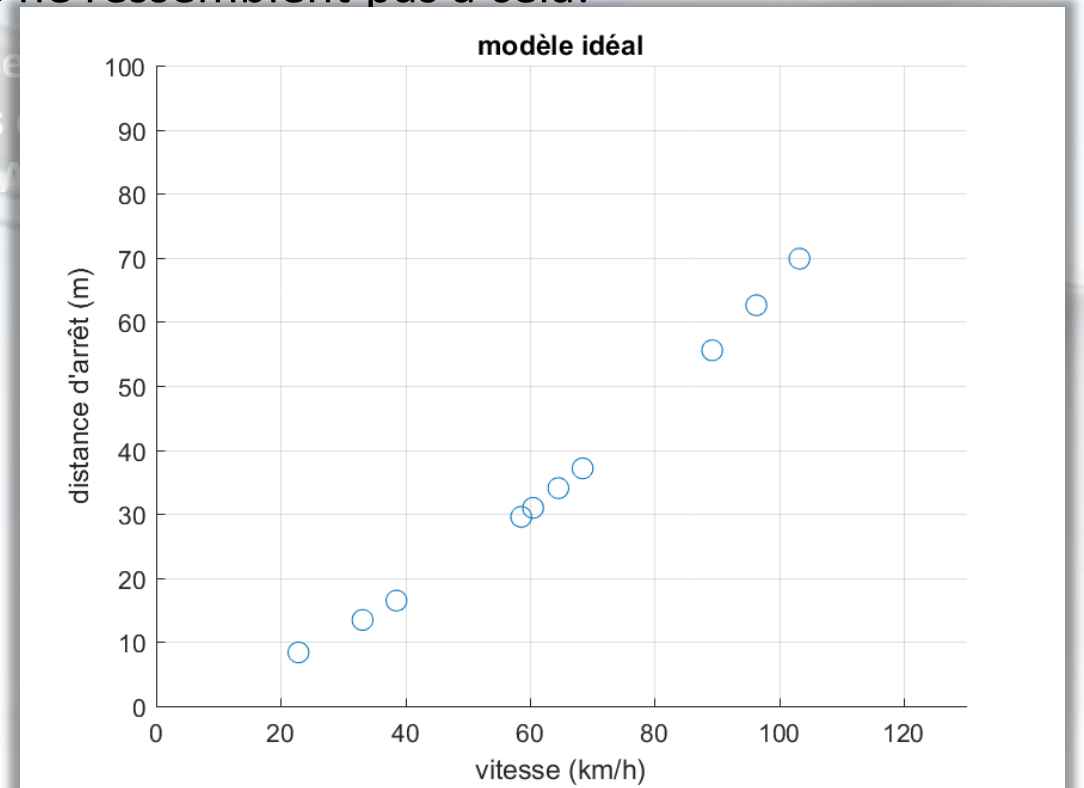
Inconvénients :

- et si aucun modèle physique satisfaisant n'est disponible ?
- et si la résolution est très coûteuse ? (temps de calcul)

Solution 3 : approche « données »

recueil des données de quelques expériences (réelles ou numériques)...

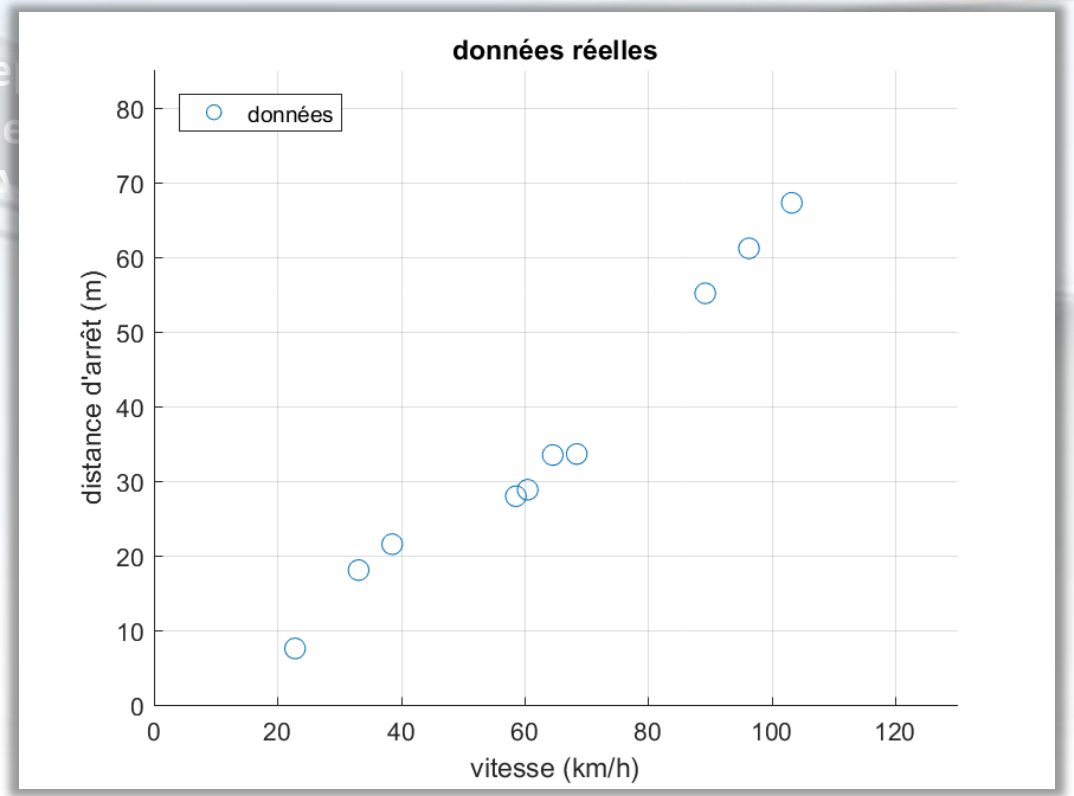
...mais les vraies données ne ressemblent pas à cela:



Solution 3 : approche « données »

recueil des données de quelques expériences (réelles ou numériques)...

… plutôt à cela:



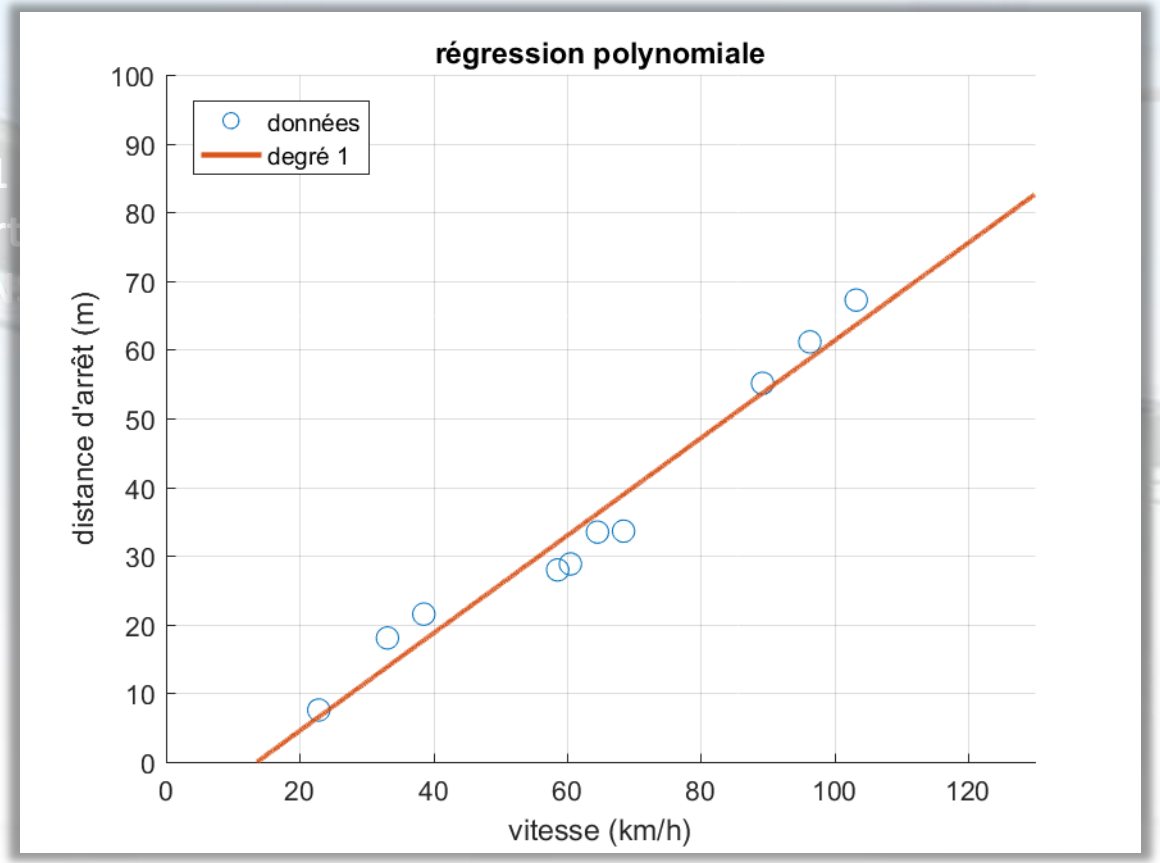
Solution 3 : approche « données »

21 septembre 2023
Arts et Métiers-Campus de Paris
ENSAM

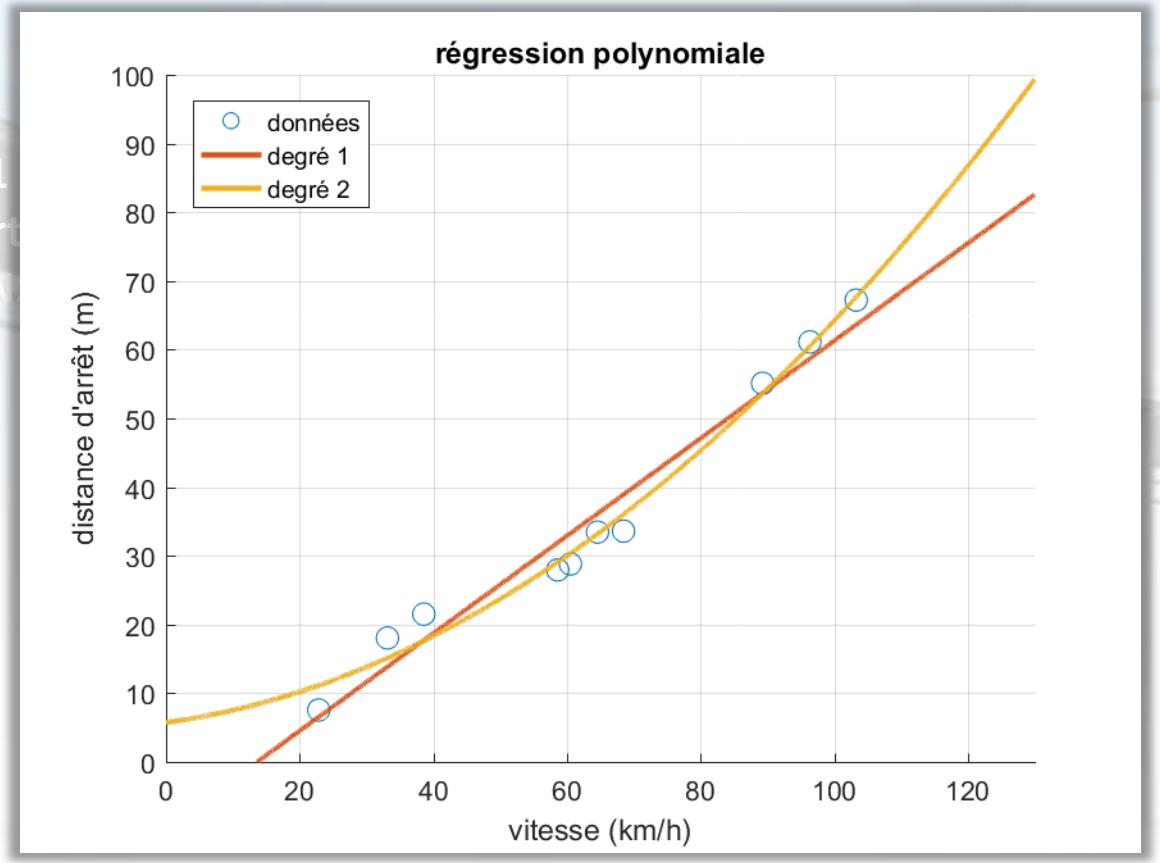
Question de l'apprentissage automatique :

quel « modèle » définir à partir des données ?

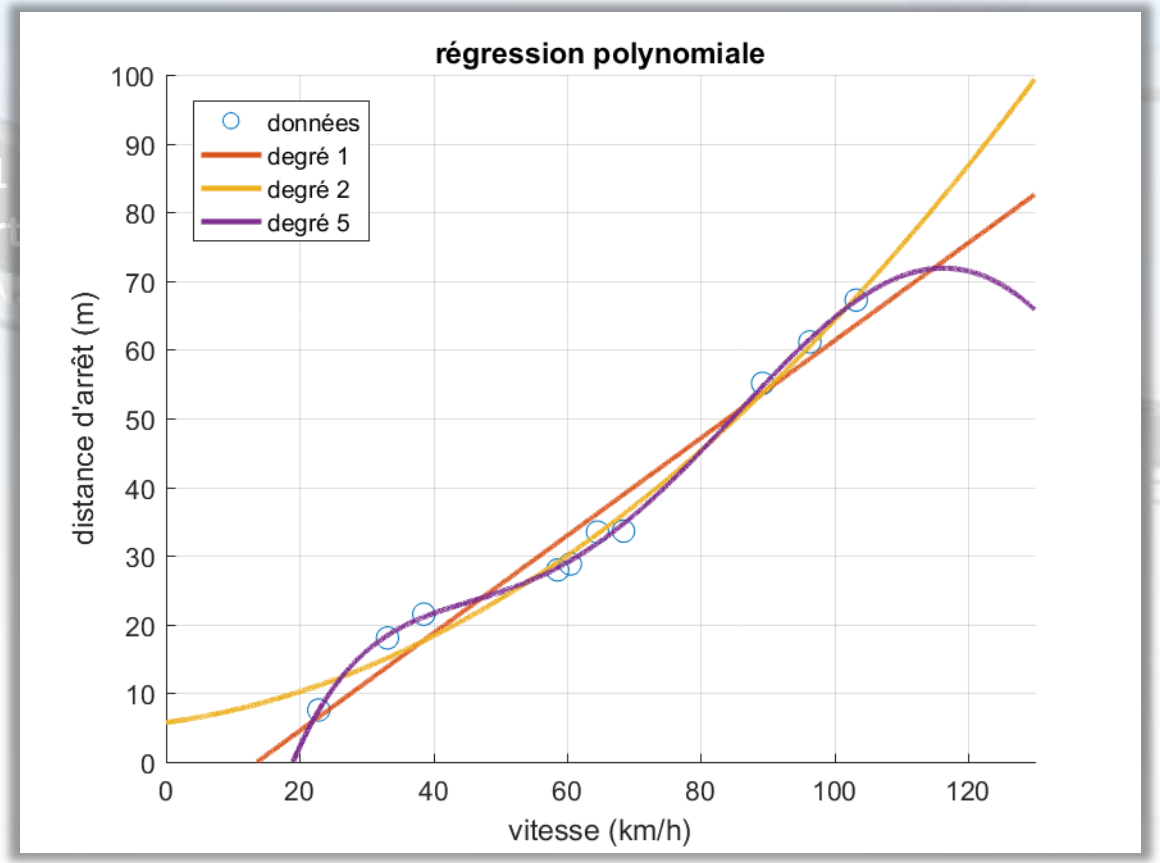
Exemple: régression



Exemple: régression



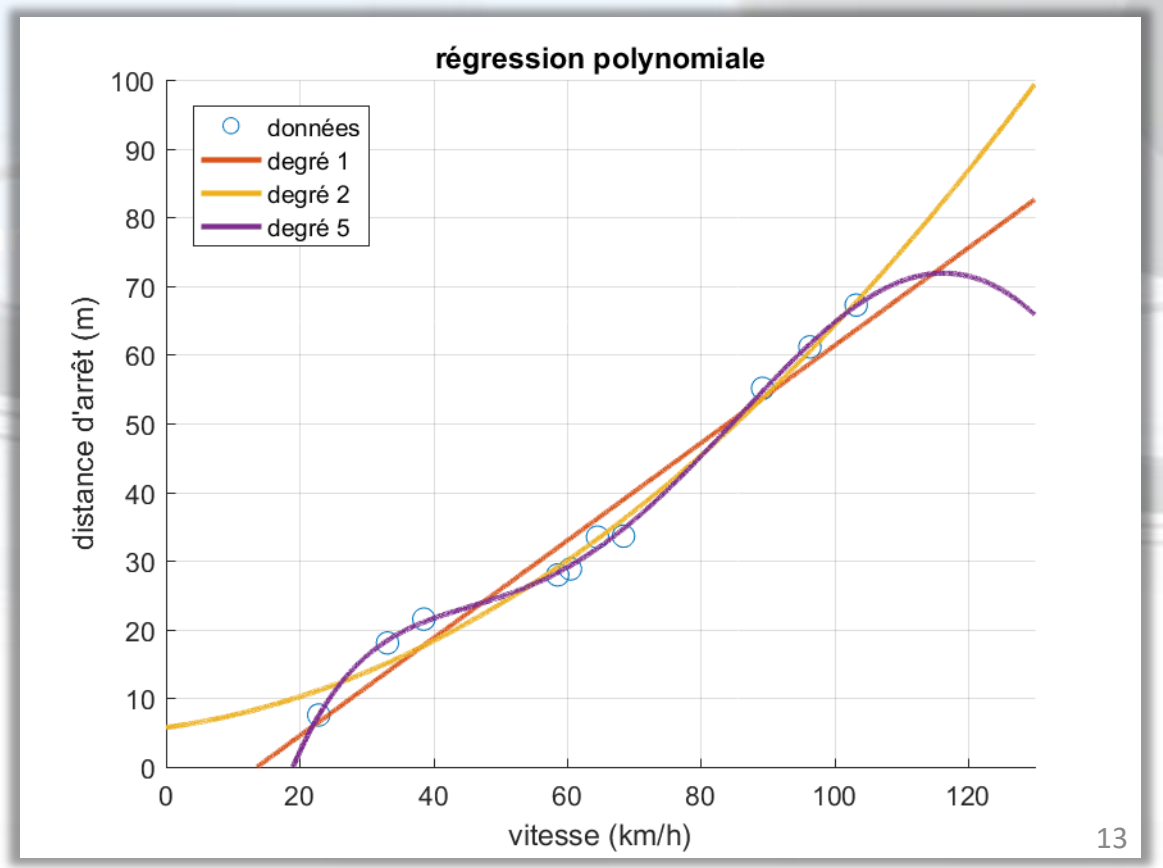
Exemple: régression



Exemple: régression

21 septembre 2023
Arts et Métiers-Campus de Pa
ENSAM

Modèle	RMSE	
Polynôme degré 1	11.2	Sous-apprentissage?
Polynôme de degré 2	7.01	
Polynôme de degré 5	2.21	Sur-apprentissage?



Objectifs du cours

- comprendre les problématiques scientifiques de l'IA
  - ➡ fondements, limites théoriques et pratiques

21 septembre 2023
Arts et Métiers-Campus de Paris
ENSAM

- comprendre ce que signifie « apprentissage »
  - ➡ principaux modèles, validation

- expérimenter dans un environnement standard utilisé en milieu professionnel / scientifique
  - ➡ carnets Jupyter et bibliothèque Python Scikit-learn

Scikit-learn " third most used free software for machine learning in the world"

But: ouvrir la boîte noire, mais cours de tronc commun, pas de maths ou info!

Le cours « Introduction à l'apprentissage automatique » à Mines Nancy

S'appuie sur **TCS 1A** : *informatique*, *probabilités*, *statistique*, *analyse numérique*, *recherche opérationnelle* et **TCS 2A** : *analyse de données*

21 septembre 2023
Arts et Métiers-Campus de Paris
ENSAM

Tout le matériel est disponible **et distribué en avance aux chargés de TD**:

- polycopié
- supports de cours
- sujets de TP (et corrections détaillées)
- compléments pour aller plus loin

https://members.loria.fr/FSur/enseignement/apprauto/
https://gitlab.univ-lorraine.fr/sur6/tcs-apprentissage-automatique-mines-nancy-2a/

TD

Question 4. Comment s'écrivent les deux modèles de prédiction de y en fonction de x? (ne reportez que quelques chiffres significatifs)

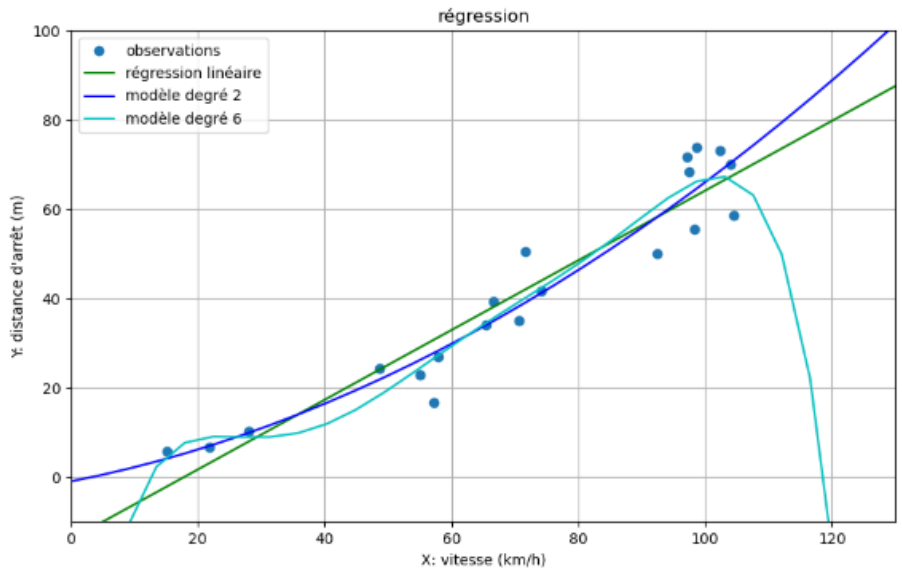
Réponse:

La représentation graphique des différents modèle est donnée par la cellule suivante.

In [7]:
```
# prédiction des deux modèles sur les valeurs X définies plus haut, pour le graphique
X2=poly2.fit_transform(X)
X6=poly6.fit_transform(X)
Y_pred_lrp2=lrp2.predict(X2)
Y_pred_lrp6=lrp6.predict(X6)

plt.figure(figsize=(10,6))
plt.plot(X_data, Y_data,'o')
plt.plot(X, Y_pred_lr, '-g')
plt.plot(X, Y_pred_lrp2, '-b')
plt.plot(X, Y_pred_lrp6, '-c')
plt.xlim(0, 130)
plt.ylim(-10, 100)
plt.xlabel("X: vitesse (km/h)")
plt.ylabel("Y: distance d'arrêt (m)")
plt.grid()
plt.title('régression')
plt.legend(["observations","régression linéaire","modèle degré 2","modèle degré 6"]);
```

Figure 3



x=64.1 y=33

Python / **scikit-learn** (tensorflow)
Jupyter notebooks: exécution interactive dans un navigateur
Une grande partie du code Python est donnée aux élèves-ingénieurs.

Exemples:
diabete, breast-cancers, wine-quality
regroupement de documents par sujets,
détection de spam,
philogénie moléculaire,
reconnaissance de chiffres manuscrits,
classification d'images…

Contenu du cours

Apprentissage supervisé (régression / classification) & **apprentissage non-supervisé** (partitionnement)

21 septembre 2023
Arts et Métiers-Campus de Paris
ENSAM

Propriétés fondamentales: malédiction de la dimension, dilemme biais-fluctuation, théorie statistique de la décision

Modèles d'apprentissage: régression ridge / lasso, k-moyennes, régression logistique, machines à vecteur support et astuce du noyau, réseaux de neurones, apprentissage profond

Responsabilité sociale et environnementale: coût énergétique de l'entraînement des gros réseaux récents, biais dans les décisions induisant des discriminations, sujets d'actualité…

Evaluation

- Un **QCM** sur Arche en début de chaque séance de TD

➡ auto-évaluation

➡ sert de base à la « note de TD » (avec participation) : 4 points /20

21 septembre 2023
Arts et Métiers-Campus de Paris
ENSAM

- Un **examen final** :

➡ vérifier que les objectifs du cours sont atteints
(pratique sur données réelles et « théorie » / méthodologie)

➡ 16 points /20

Difficultés

- Peu de temps, rythme intense

21 septembre 2023
Arts et Métiers-Campus de Paris
ENSAM

- Appui fort sur le tronc commun de maths / info: quid des AST?

- Hétérogénéité des attentes:

matheux / geeks / applications / poursuite en M2 / profil recherche…

- Difficultés en informatique « de base »!

Conclusion

Le cours « Introduction à l'apprentissage automatique » à Mines Nancy…

… essaie d'équilibrer **pratique** et **théorie**, met l'accent sur la **pluridisciplinarité**

21 septembre 2023
Arts et Métiers-Campus de Paris
ENSAM

… permet de:

dépasser l'usage de « l'IA » en boîte noire, pouvoir décider de « solutions IA »

mieux appréhender les **projets scientifiques** et **industriels**, **stages** en entreprise…

armer les étudiants aux **poursuites d'études** dans le domaine

Pistes d'ouverture: génération d'images (Stable diffusion & co), de textes (chatGPT) ?