

Modélisation stratégies de pêche

Réunion du 18/12/2019

S. Moinard, T. Le Minh

Réunion précédente

(25/11/2019)

▶ Avancées :

- ▶ Clusters : Perte d'info en sommant les quantités
- ▶ Sélection de variables (Lasso, AIC stepwise)

▶ Idées :

- ▶ Stratégie par strate : détecter les zéros, puis régression
- ▶ Influence de la géographie : rectangles, interactions avec mois et espèces
- ▶ Modèles : non-probabilistes (random forests), graphiques (GGM)

Notations

- ▶ $n=362$ observations (opérations de pêches)
- ▶ Données sur $K=150$ espèces, indexées par k
- ▶ Variables prédictives :
 - ▶ Pour chaque k , X_k = donnée de débarquement de l'espèce k
 - ▶ month = mois de l'année
 - ▶ rect = nom du rectangle
- ▶ Variables à prédire :
 - ▶ Pour chaque k , Y_k = donnée de rejet de l'espèce k

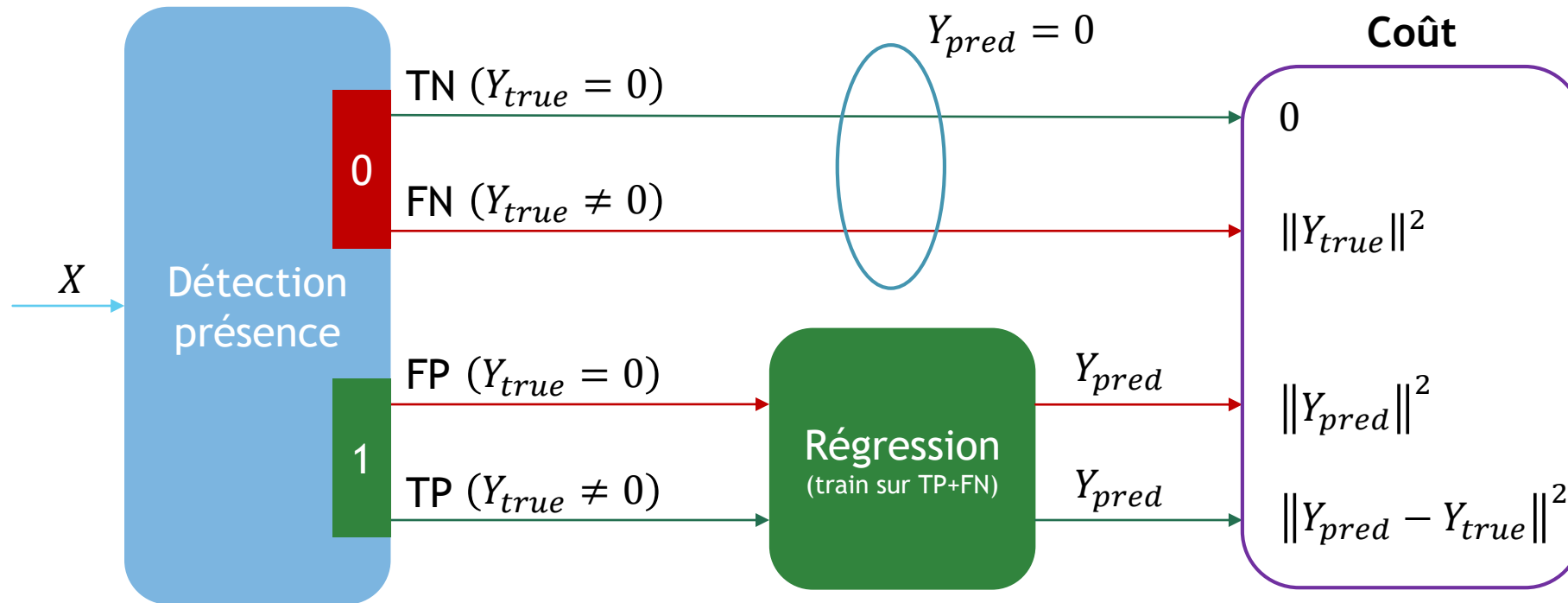
Détection de zéros

- ▶ $X_k, Y_k \in \{0,1\}$, absence/présence dans le débarquement/rejet
- ▶ Classification binaire
 - ▶ Régression logistique
 - ▶ Un modèle par espèce Y_k
- ▶ Analyse des taux de FN et FP
 - ▶ Réglage du seuil de décision

Régression

- ▶ $X_k, Y_k \in \mathbb{R}$, quantité débarquée/rejetée
- ▶ Réglage des paramètres avec ensemble de train/test
 - ▶ Paramètres différents pour chaque Y_k
- ▶ Modèles
 - ▶ Modèles linéaires + sélection de variables
 - ▶ Arbre de décision : surajusté
 - ▶ Random forest

Procédure (NZ)



Amélioration des résultats

- ▶ Réglages des paramètres modèles
- ▶ Sélection de variables
 - ▶ Lasso
 - ▶ Step AIC / modèles graphiques
- ▶ Regroupement des variables sélectionnées
 - ▶ Clustering sur les X_k
- ▶ Regroupement des paramètres pour la régression
 - ▶ Clustering sur les Y_k

Axes d'exploration

- ▶ Modèles graphiques (GGM, GGM+copules)
 - ▶ Problème des zéros
 - ▶ Autres distributions ?
 - ▶ Lois log-normales
 - ▶ Comment sélectionner un graphe ?
- ▶ Clustering + modèle graphique ?
 - ▶ Modèle à blocs latents ?