Cours 20: Séries Temporelles

IFT 6758

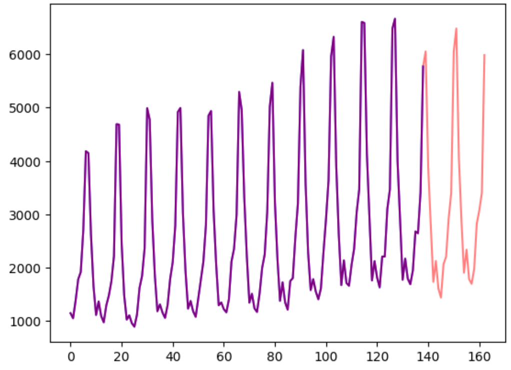
Université de Montréal

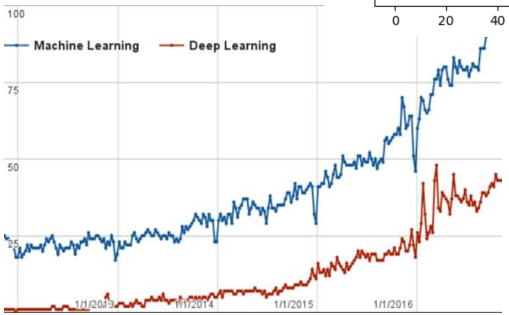
Mila

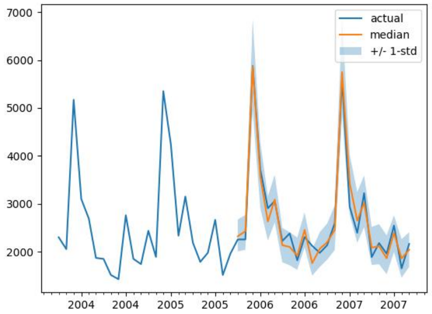
DIRO

Motivation

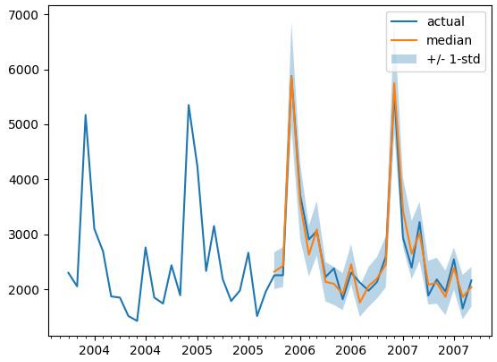
- Exemples de séries temporelles
  - Valeurs boursières
  - Variables économiques
  - Température (météo)
  - Tremblement de terre
  - Demande/consomation d’énergie



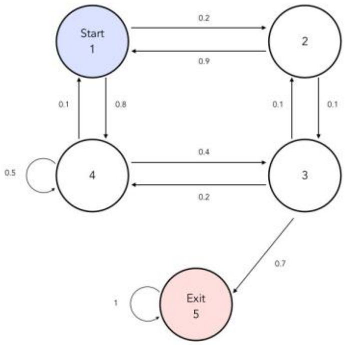


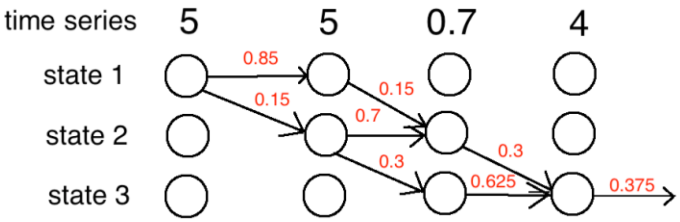


Méthodes Auto-Regressives



- La variable au temps t+1 est une fonction de x_t
  - $X_{t+1} = f(x_t, w)$
  - But: apprendre les paramètres w du modèle pour prédire le futur.
- Utilise l'hypothèse de Markov
  - L'état x_t contient assez d'information pour prédire l'état suivant.





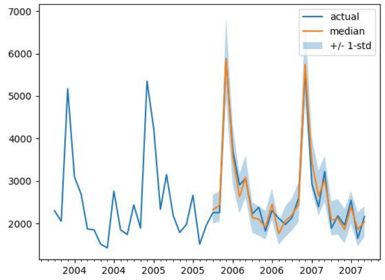
Méthodes Autoregressives

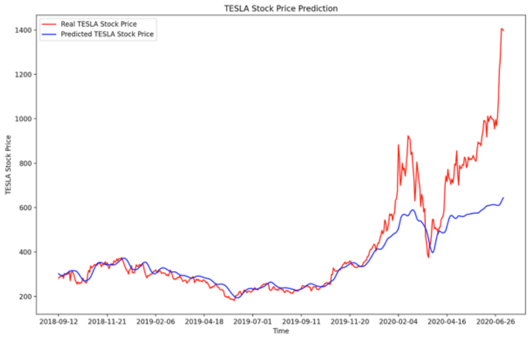
- Modèles
  - Il est possible d’utiliser de nombreux modèles vu en classe pour modèliser la fonction f.
  - Modèles linéaires, réseaux de neurones, tranformers

- Hypothèse de Markov d’ordre q (vu en language)
  - Les états $x_{t-q+1},\ldots,x_t$ sont suffisants pour prédire x_{t+1}
  - Une certaine quantité de l’historique est suffisant pour prédure le futur.

À quel point à-t-on besoin du passé?

- *q* est une hyperparamètre du modèle
  - Il va dépendre du jeu de donné et de l'application
  - Plus q est grand, plus le modèle a de paramètres
  - Important d'avoir le bon ordre de grandeur.
- Exemples
  - Prédire la météo: (de 2 jours à 200 ans)
  - Consomation énergétique (cycles journaliers et hebdomadaires)
  - Actions boursières (cycles trimestriels).





Apprentissage hors ligne (offline learning)

Méthodes d'apprentissage

- *Modèles linéaires autoregressifs*

$$Y_t = \gamma + \sum_{i=1}^{q} a_i Y_{t-i} + \epsilon_t$$

Avec:

- \epsilon_t est le bruit dans les mesures.
- a_0, $a_1, \ldots, a_q$ sont les coefficients autoregressifs.
- Y_t est la variable observée.

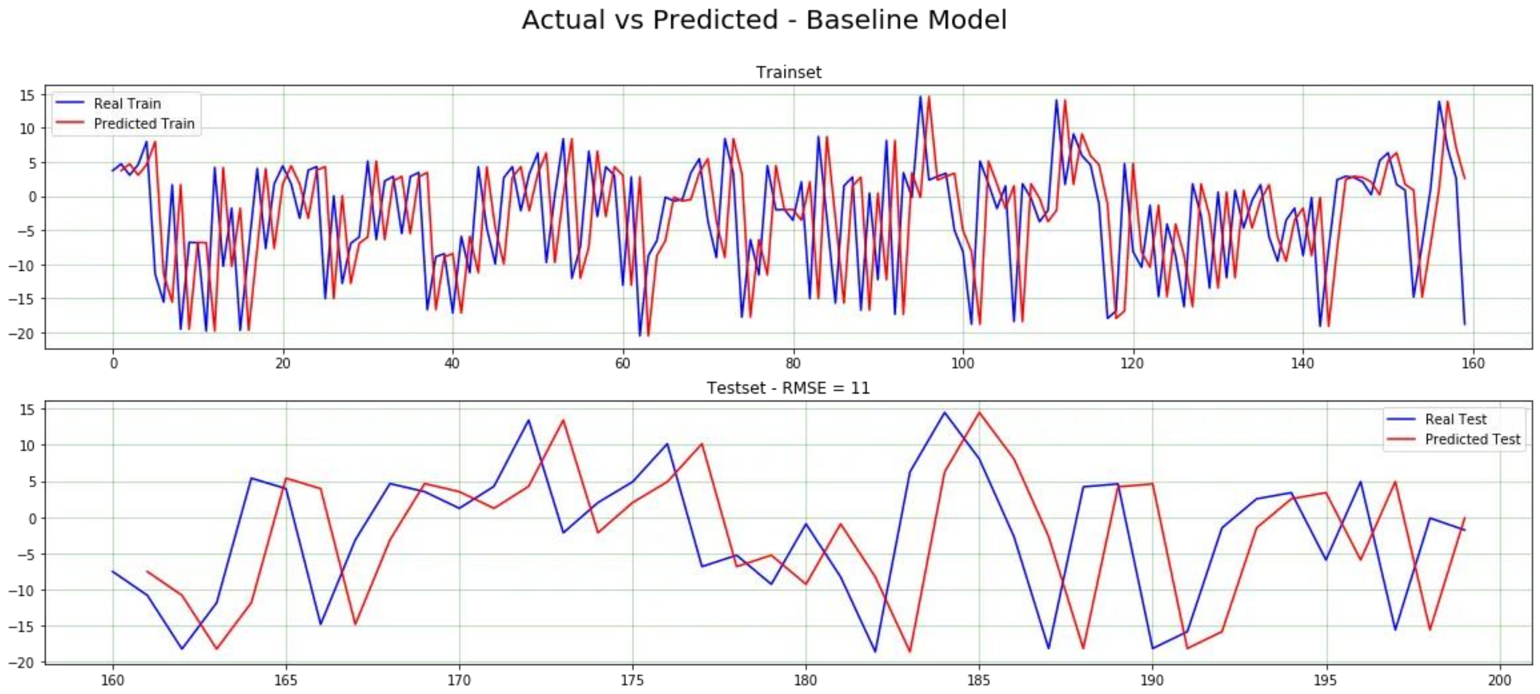
Exemple:

Caractéristiques:
$Y_{t-1}, \ldots, Y_{t-q}$

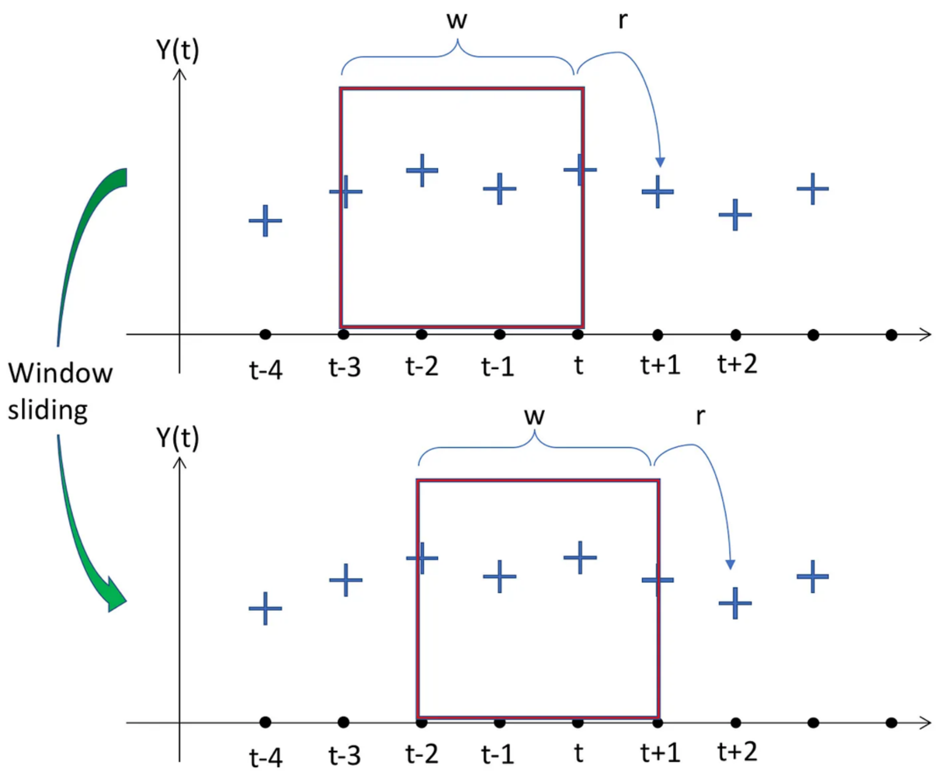
Étiquette:
Y_t

Paramètres:
a_i

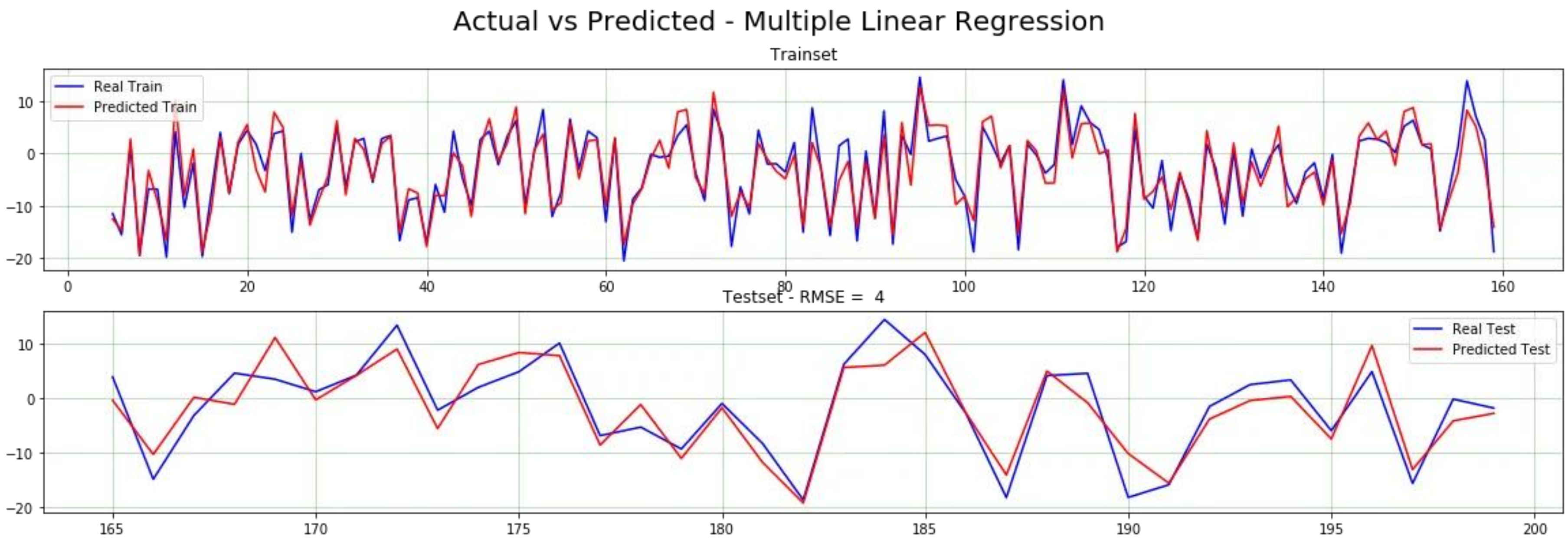
Le modèle le plus simple: $\hat{y}_t = y_{t-1}$



Exemple



Modèle linéaire autoregressif



Encore mieux: ARMA et ARIMA

- ARMA est la combinaison d’un modèle Auto-Regressive et d’un modèle (MA) Moving Average:

$$Y_t = \gamma + \sum_{i=1}^{q} a_i Y_{t-i} + b_i \epsilon_{t-i} + \epsilon_t$$

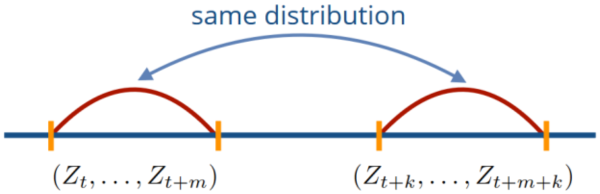
ARMA et ARIMA

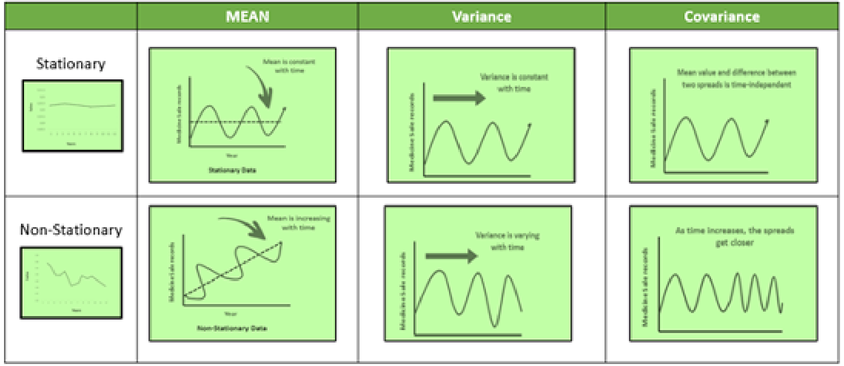
$$Y_t = \gamma + \sum_{i=1}^{q} a_i Y_{t-i} + b_i \epsilon_{t-i} + \epsilon_t$$

- ARMA is est idéal pour les séries stationnaires. **ARIMA** a été developpé pour fonctionner sur les séries non-stationnaires.
  - AR ➔ Utilise les valeurs passées pour prédire le futur.
  - MA ➔ Utilise les erreurs passées pour prédire l’erreur future.
  - I==> Transformation de l’entrée Y pour la rendre plus stationnaire.

Stationnarité d’une série temporelle

Définition : une séquence de variables aléatoires est stationnaire si sa distribution est invariante au décalage temporel.





Designed by Author (Shanthababu)

Modèles de Markov avec variables cachées

- *La non-stationnarité apparait lorsque l'on manque d'information*
- Nous observons seulement l'état Y.

$$\mathbf{X}_{t+1} = \mathbf{B}\mathbf{X}_t + \mathbf{U}_t,$$
$$Y_t = \mathbf{A}\mathbf{X}_t + \epsilon_t$$

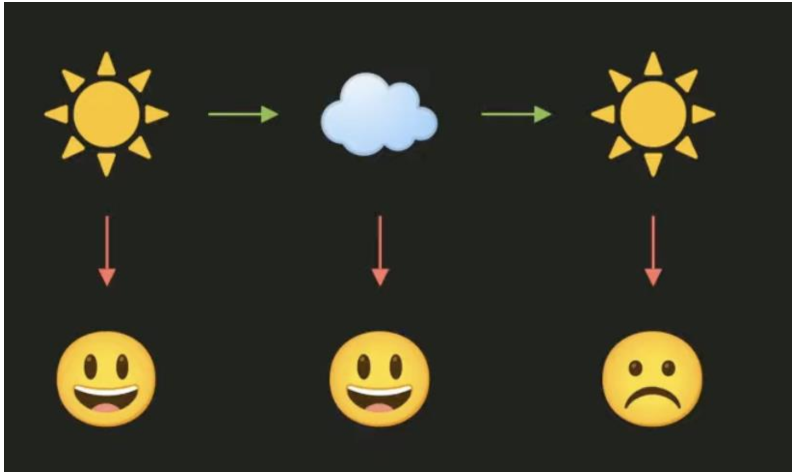
  - $\mathbf{X}_t$ is an n-dimensional state vector.
  - Y_t is an observed stochastic process.
  - $\mathbf{A}$ and $\mathbf{B}$ are model parameters.
  - $\mathbf{U}_t$ and ϵ_t are noise terms.

Modèles de Markov avec variables cachées

- *La non-stationnarité apparait souvent lorsque l'on manque d'information*

Variable Cachée:
(imaginons que nous n'avons pas accès à la météo)

Variable Observée:



Modèles de Markov avec variables cachées

- *La non-stationnarité apparait lorsque l'on manque d'information*
- Nous observons seulement l'état Y.

$$\mathbf{X}_{t+1} = \mathbf{B}\mathbf{X}_t + \mathbf{U}_t,$$
$$Y_t = \mathbf{A}\mathbf{X}_t + \epsilon_t$$

- La pluspart du temps:
  - On ne peut pas observer X_t .
  - On sait qu'il existe (donc on va essayer de l'apprendre)
  - Principe des RNN (Réseaux neuronaux récurrents) (Les RNNs sont **Non linéaires)**

Généralement, deux scénarios

Scénario Stochastique

- Hypothèse que la distribution est stationnaire (ou avec une non-stationnarité simple)
- Mesure de performace: perte moyenne

Scénario en-ligne:

- Pas d'hypothèse sur la distribution
- Mesure de performance: regret.
- Recherche active: (Cesa-Bianchi and Lugosi, 2006; Anava et al.
- 2013, 2015, 2016; Bousquet and Warmuth, 2002; Herbster and Warmuth,
- 1998, 2001; Koolen et al., 2015).

Apprentissage en-ligne (online learning)

Apprentissage en-ligne

- Apprentissage en ligne:
  - On recoit l'entrée x_t
  - On choisit un classifieur h_t
  - L'environnement choisit y_t
  - On souffre de la perte L
- Objectif:
  - Minimiser le regret.
  - On ne veut plus minimiser la perte mais la somme des pertes.

- For $t = 1$ to T do
  - player receives $x_t \in \mathcal{X}$.
  - player selects $h_t \in H$.
  - adversary selects $y_t \in \mathcal{Y}$.
  - player incurs loss $L(h_t(x_t), y_t)$.
- Objective: minimize (external) regret

$$\mathrm{Reg}_T = \sum_{t=1}^{T} L(h_t(x_t), y_t) - \min_{h \in H^*} \sum_{t=1}^{T} L(h(x_t), y_t).$$

Apprentissage en-ligne

- Apprentissage en ligne:
  - On recoit l’entrée x_t
  - On choisit un classifieur h_t
  - L’environnement choisit y_t
  - On souffre de la perte L

$$\mathrm{Reg}_T = \sum_{t=1}^{T} L(h_t(x_t), y_t) - \min_{h \in H^*} \sum_{t=1}^{T} L(h(x_t), y_t).$$

- Résumé:
  - L’environnement a tout les pouvoirs.
  - On ne peut plus minimiser la perte.
  - On peut minimiser le regret (si la tâche était trop dure alors on a rien a regretter)
- En pratique:
  - **Risqué car potentiellement instable avec de l’apprentissage profond.**
    (sauf avec des modèles simples qui ont des guaranties)

Peut-on trouver un juste milieu?

- Apprentisage hors ligne (sur l'ensemble des données passées) est plus stable
  - Ne capture pas les changement de distributrion
  - Peut-on le combiner avec l'apprentissage en ligne?
- Oui: Minimisation répétée
  - Entraîner sur les données passées
  - Évaluer le modèle
  - Deployer le modèle
  - Collecter de nouvelles données (ne pas les utiliser en ligne)
  - Répéter

Deux exemples en pratique

- Prediction performative

- Adapation aux données test sans étiquettes.

Apprentissage de données statiques

Mais dans de nombreuses applications, la distribution va changer. En particulier les nouvelles données peuvent dépendre du modèle deployé.

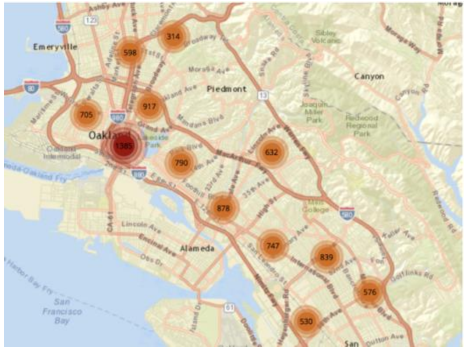
Fermer la boucle!

Data Distribution —*Observations*→ *Predictor*

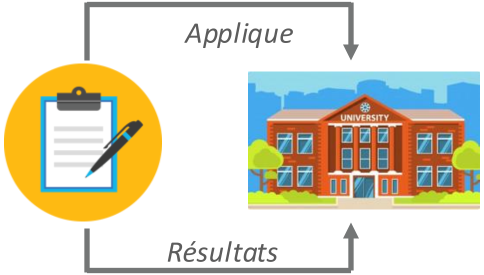
Decision

Traditionellement ignoré

Performativité: La distribution des données dépend du modèle



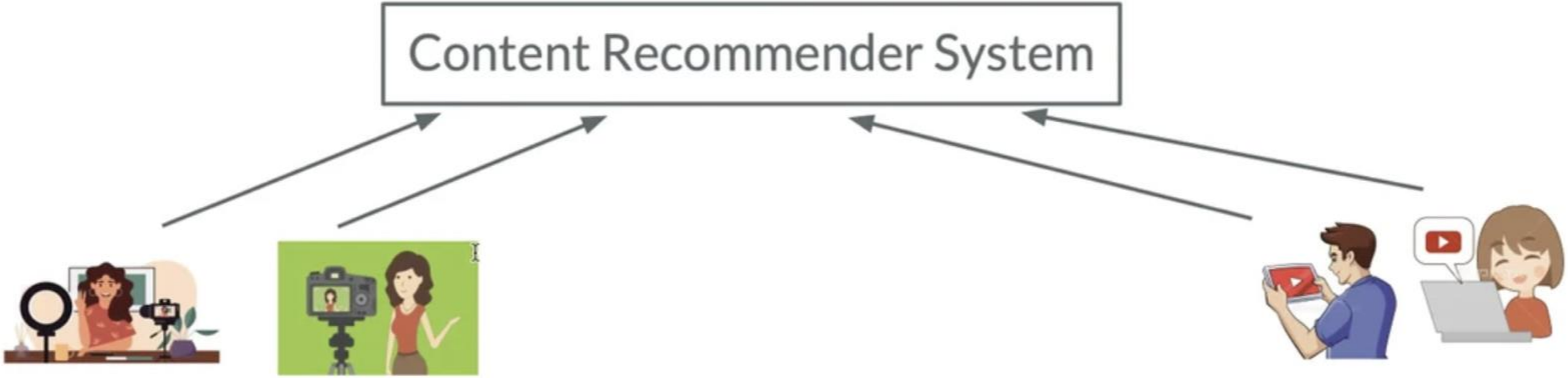
Prédiction de crime



Admission

Social Impact of Algorithmic Decisions

Algorithmic decisions can *shape* the marketplace where they are deployed.



Minimisation du risque répétée:

- La distribution à l'étape t dépend du classifieur déployé au temps précédent

$$\theta_{t+1} = \arg\min_{\theta} \mathbf{E}_{(x,y)\sim p_{\theta_t}} [\ell(f_\theta(x), y)]$$

- La performance après déploiement est différente de la performance mesurée avant le déploiment:

$$\mathbb{E}_{(x,y)\sim p_{\theta_t}} [\ell(f_{\theta_t}(x), y)] \neq \mathbb{E}_{(x,y)\sim p_{\theta_{t+1}}} [\ell(f_{\theta_t}(x), y)]$$

Equilibre

- Lorsque le modèle induit la distribution sur laquelle il a été appris on a atteint la stabilité.

$$\theta_* = \arg\min_\theta \mathbb{E}_{(x,y)\sim p_{\theta_*}} [\ell(f_\theta(x), y)]$$

- En pratique: Important de monitorer le modèle pour détecter les baisses de performance (et réentrainer!)

Adaptation sans étiquettes

- Vous entraîner votre modèle sur un jeu de données classique
- Vous deployer votre modèle sur un tâche réelle.
- Des données arrivent aprés le deploiement.
- Elles correspondent parfaitement à la tâche!
- Problème: **pas d'étiquettes**

- Comment les utiliser ?

Idée: générer les étiquettes avec le modèle

- Pseudo étiquettes: $\hat{y} = f_\theta(x^t)$.
- Ne marche que si le modèle est bon à la base!

Published as a conference paper at ICLR 2021

TENT: FULLY TEST-TIME ADAPTATION BY ENTROPY MINIMIZATION

Dequan Wang[1]*, Evan Shelhamer[2*†], Shaoteng Liu[1], Bruno Olshausen[1], Trevor Darrell[1]
dqwang@cs.berkeley.edu, shelhamer@google.com
UC Berkeley[1] Adobe Research[2]

Conclusion

- Series temporelles:
  - Utiliser l’hypothèse de Markov!
  - Les valeurs du passé sont utiles pour prédire le futur
  - Important de vérifier que la distribution est stationnaire!

- Non-stationnarité de la distribution
  - Peut être du à
    - Changement saisonaux
    - Performativité
  - Important de monitorer les performances des modèles deployé
  - Ré-entrainer (finetune) quand nécessaire.
  - On peut utiliser le modèle courant pour générer des pseudo-labels (attention risqué)

References

- https://www.taylorfrancis.com/books/mono/10.1201/9781420010893/hidden-markov-models-time-series-walter-zucchini-iain-macdonald
- https://cims.nyu.edu/~mohri/amls/aml_time_series.pdf
- https://www.slideshare.net/DerekKane/data-science-part-x-time-series-forecasting
- https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-to-time-series-analysis/