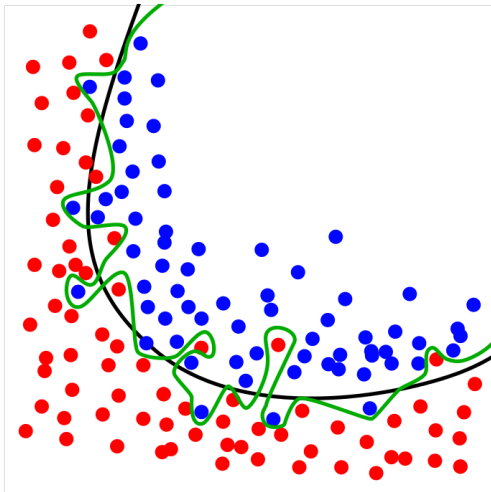


Problématiques liées au machine learning

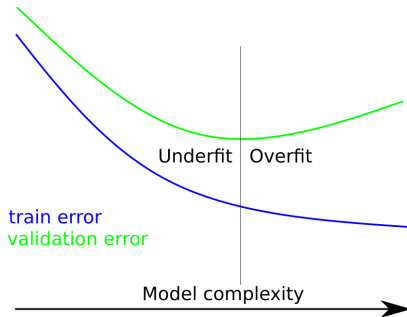
Fin de la magie et retour à la réalité

- Les données (trouées, incomplètes, non-représentatives, biais statistiques,...)
- Une fonction de perte pas adaptée au besoin
- La reproductibilité (random seed)
- La fuite d'information dans le modèle
- La quasi non-interprétation du modèle
- Le Surapprentissage

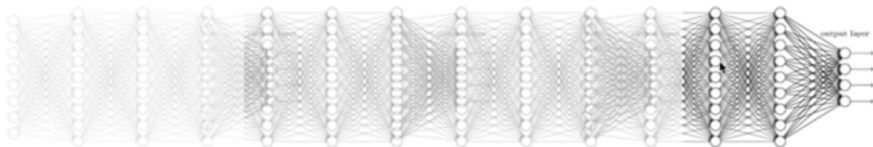
Overfitting



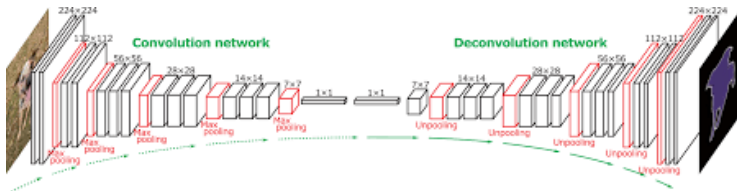
Les Problèmes théoriques



Disparition du gradient



Disparition du gradient



Les **données** coûtent cher (récolte, nettoyage)



Les Problèmes logistiques

L'apprentissage d'un modèle prend **beaucoup** de **temps**

- Reconnaissance parole : 4 GPU => plusieurs jours, semaines
- Alphago : 3 semaines sur 5000 TPU (\approx 30M\$)
- On n'obtient pas de résultats concluant avec le premier run.
- Peu de visibilité sur le temps d'obtention d'une plus value (en fonction du domaine)

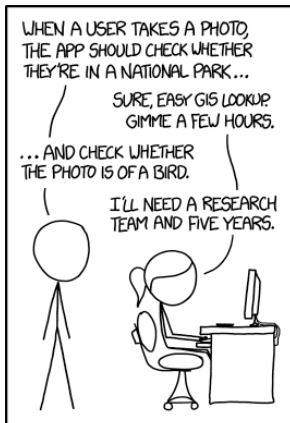


Les Problèmes logistiques

L'utilisation de **GPU/TPU** coûte cher
...et ça ne s'arrange pas avec l'explosion du besoin par le minage de cryptomonnaies



Cours de l'action NVIDIA



IN CS, IT CAN BE HARD TO EXPLAIN
THE DIFFERENCE BETWEEN THE EASY
AND THE VIRTUALLY IMPOSSIBLE.

Les Problèmes logistiques

Les **ingénieurs en machine learning** (compétents) coûtent cher :

- git clone d'un papier + jouer avec data \neq data-scientist
- au croisement de l'ingénierie (traitement de gros volumes de données, standard de développement) et des mathématiques (statistiques, algèbre, optimisation)
- TOUS les meilleurs sont/vont dans 4 entreprises ! (ou presque...)



Les Problèmes de communication

Utilisateurs prêts à accepter un algorithme qui fait des erreurs ?



Les Problèmes de communication

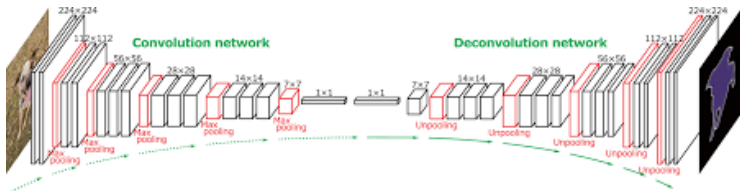
Utilisateurs prêts à accepter un algorithme qui fait des erreurs ?



ML WEEK

<https://www.ml-week.com/> À l'usage exclusif des participants aux formations

Utilisateurs prêts à accepter un algorithme qui n'est pas interprétable ?



Les Problèmes Ethiques

Grande-Bretagne : Predpol (abandonné), National Data Analytics Solution (NDAS) en mars 2019



Les Problèmes Ethiques

Détection mensonge à la frontière (iBorderCtrl)

Projet pilote en Hongrie, Lettonie et Grèce



Les Problèmes Ethiques

Chine 170 Millions de caméras (450 Millions en 2020)



Les Problèmes Ethiques

Chine 170 Millions de caméras (450 en 2020)

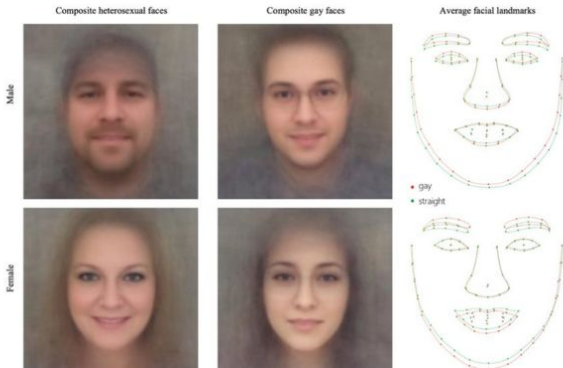


Les Problèmes Ethiques

Automatisation du processus de recrutement



Préférence sexuelles à partir d'une photo ?



Évaluer une Baby-sitter ?



Robots tueurs ?

