

Préparation Quizz 2

OPT6 : Apprentissage Avancé

Les castors juniors

December 2017

1 Questions

Dans quel cas le principe de minimisation du risque empirique est-il sain ?

S'il y a des contraintes sur l'espace des hypothèses.

Qu'est-ce qu'une hypothèse de risque empirique nul ?

Pas d'erreur sur l'échantillon d'apprentissage.

Qu'est-ce qu'un cas réalisable ?

$\exists h \in H$ tq $R(h) = 0$ (Risque réel, pas empirique) Avec le perceptron ça veut dire qu'il va converger surtout. Si y'a pas de cas réalisable, le perceptron convergera pas.

Qu'est-ce que iid ?

Indépendemment identiquement distribué.

Each random variable has the same probability distribution as the others and all are mutually independent. Sur un ensemble de test ça permet d'être sûr que l'ensemble est bien représentatif des données réelles (?)

Le risque empirique régularisé ?

C'est un critère inductif qui consiste à d'abord Satisfaire les contraintes posées par les exemples, puis à Choisir le meilleur espace d'hypothèses (capacité de H)

La dimension Vapnik-Chervonenkis ?

La dimension VC (pour dimension de Vapnik-Chervonenkis) est une mesure de la capacité d'un algorithme de classification statistique. Elle est définie comme le cardinal du plus grand ensemble de points que l'algorithme peut pulvériser.

C'est une sorte de mesure de la complexité d'un modèle.

- Séparation par une droite : $D_{VC} = 3$
- Séparation par un rectangle : $D_{VC} = 4$ (je crois)

Pulvérisation ?

Soient C une classe d'ensembles et A un ensemble. On dit que C pulvérise A si et seulement si, pour tout sous-ensemble T de A , il existe un élément U appartenant à C tel que

$$U \cap A = T$$

Ceci équivaut encore à dire que C pulvérise A si et seulement si l'ensemble des parties de l'ensemble A , $P(A)$, est égal à l'ensemble $U \cap A | U \in C$.

Par exemple, la classe C des disques du plan (lorsqu'on se place dans un espace à deux dimensions) ne peut pas pulvériser tous les ensembles F de quatre points, alors qu'en revanche la classe des ensembles convexes du plan pulvérise tout ensemble fini du cercle unité.

Apprenant faible / apprenant fort ?

- **Apprenant faible** : Qui fait un tout petit peu mieux que la moyenne
- **Apprenant fort** : PAC learning

Y a-t-il un rapport entre les deux ?

Oui, on peut faire un apprenant fort avec plein d'apprenants faibles (== boosting)

PAC Learning ?

Probably Approximately Correct.

Qu'est-ce qu'un decision stump ?

Arbre de décision à un seul noeud. Typiquement un apprenant faible. Découpe l'espace en deux selon un des attributs.

Comment engendrer des apprenants faibles décorrélés ?

En modifiant l'échantillon d'apprentissage à chaque étape : on diminue l'importance des exemples bien classés et on augmente celle des exemples mal classés.

Consistance ?

Qu'est-ce que le principe de consistance universelle ?

Les résultats de la consistance universelle sont des résultats asymptotiques. Ils disent juste que si nous avons suffisamment de données (et un ordinateur suffisamment puissant pour les traiter) alors tout algorithme universellement consistant produira une prédiction très proche de la meilleure prédiction possible.

La loi des grands nombres (Law of Large Numbers) ?

Exprime le fait que les caractéristiques d'un échantillon aléatoire se rapprochent des caractéristiques statistiques de la population lorsque la taille de l'échantillon augmente.

Qu'est-ce que le boosting ?

Une méthode générale pour convertir des règles de prédiction peu performantes en une règle de prédiction très performante.

Luckiness Framework ?

Principe : définir un ordre sur H qui dépend des données. Si nous avons de la chance alors il n'y aura pas trop d'hypothèses mauvaises aussi compatibles avec la cible que les bonnes.

La sagesse des foules (wisdom of crowd) ?

Le fait d'élire Sarko puis Hollande puis Macron. C'est pas plutôt genre les paniers de bouffe dans les marchés ? Ah oui peut être. Les gens sont cons mais tous ensemble leur moyenne donne l'espérance d'une loi normale avec une bonne précision.

2 Formules

Pour l'algorithme du boosting :

Soit e le taux d'erreur d'une hypothèse h :

- $\alpha = \frac{1}{2} \ln\left(\frac{1-e}{e}\right)$ (c'est le coeff de l'hypothèse dans le vote à la fin)
- $P_b(x) = \frac{1}{(2(1-e))}$ (on multiplie l'ancien coeff des valeurs bien prédites par ça pour avoir leurs nouveaux coeffs)
- $P_m(x) = \frac{1}{2e}$ (on multiplie l'ancien coeff des valeurs mal prédites par ça pour avoir leurs nouveaux coeffs)