
Mesures d'importance basées sur la variance dans le contexte de la régression linéaire : analyses comparatives et tests numériques.

Vincent Chabridon*¹

¹EDF, Chatou – EDF, Chatou – France

Résumé

En apprentissage statistique supervisé, une des questions les plus couramment rencontrées dans de nombreuses études (industrielles ou de recherche) est l'identification des variables influentes (en entrée) liées aux phénomènes d'intérêt observés ou mesurés en sortie. Dans le contexte de la régression linéaire, les mesures d'importance sont des outils qui ont su démontrer leur efficacité pour la sélection des régresseurs (ou prédicteurs) et l'interprétation des modèles, car elles permettent d'identifier et de classer les régresseurs les plus influents. En particulier, les mesures d'importance basées sur la variance ont été (et sont toujours) un domaine de recherche intense dans le domaine des statistiques, ainsi que dans le domaine connexe de l'analyse de sensibilité globale pour les expériences numériques. Ceci est dû à leur capacité interprétative en tant que "parts de variance de la variable expliquée". Cet exposé se concentrera donc sur le modèle de régression linéaire et visera à fournir une vue d'ensemble actualisée des méthodes de construction de mesures d'importance basées sur la décomposition de la variance, et ce, vis-à-vis, de deux verrous principaux que sont : d'une part, le cas des entrées dépendantes; et d'autre part, le cas de la grande dimension des entrées. Enfin, une illustration de la mise en oeuvre numérique sur des données simulées, publiques et industrielles est proposée.

*Intervenant