

Régression Linéaire vs Support Vector Regression (SVR)

Youssef SALMAN

1 Introduction

La régression linéaire et la Support Vector Regression (SVR) sont deux méthodes de prédiction d'une variable réelle. Elles ont des points communs, mais reposent sur des principes théoriques très différents. Ce document résume leurs différences essentielles.

2 Forme des modèles

2.1 Régression linéaire

La régression linéaire cherche une relation affine entre les données :

$$\hat{y}(x) = \beta_0 + \boldsymbol{\beta}^\top x.$$

Les coefficients minimisent la somme des erreurs quadratiques :

$$\min_{\beta_0, \boldsymbol{\beta}} \sum_{i=1}^n (y_i - \beta_0 - \boldsymbol{\beta}^\top x_i)^2.$$

2.2 SVR linéaire

La SVR cherche une fonction plate qui reste dans un **tube** d'épaisseur ε autour des données :

$$|y_i - \hat{f}(x_i)| \leq \varepsilon.$$

La forme finale du modèle est :

$$\hat{f}(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b,$$

où seuls les vecteurs de support vérifient

$$|\alpha_i - \alpha_i^*| > 0.$$

2.3 SVR non-linéaire (avec noyau)

Grâce aux noyaux :

$$K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle,$$

la SVR devient non linéaire :

$$\hat{f}(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b.$$

3 Objectifs d'optimisation

3.1 Régression linéaire

$$\min_{\beta_0, \beta} \sum_{i=1}^n (y_i - \hat{y}(x_i))^2.$$

- Toutes les erreurs sont pénalisées.
- Les valeurs aberrantes ont une grande influence.
- Pas de contrôle explicite de la complexité.

3.2 SVR

$$\min_{\beta, b, \xi_i, \xi_i^*} \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*).$$

$$\begin{cases} y_i - \langle \beta, x_i \rangle - b \leq \varepsilon + \xi_i, \\ \langle \beta, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^*. \end{cases}$$

- Les petites erreurs (dans le tube) sont ignorées.
- Les points sur ou en dehors du tube deviennent des vecteurs de support.
- La complexité est explicitement contrôlée par $\|\beta\|^2$.
- Le paramètre C règle la pénalisation des erreurs.

4 Comparaison synthétique

	Régression linéaire	SVR
Sortie	réelle	réelle
Fonction objectif	$\sum (y_i - \hat{y}_i)^2$	$\frac{1}{2} \ \beta\ ^2 + C(\xi_i + \xi_i^*)$
Traitement petites erreurs	toutes pénalisées	ignorées si $\leq \varepsilon$
Effet des outliers	sensible	robuste
Non-linéarité	doit être ajoutée manuellement	naturelle via les noyaux
Points influents	tous les points	uniquement les vecteurs de support

5 Conclusion

La régression linéaire est simple et rapide, mais sensible au bruit et limité au cadre linéaire. La SVR offre un contrôle de la complexité, ignore les petites erreurs, est robuste et permet des modèles non linéaires grâce aux noyaux. Elle est donc plus flexible et plus puissante dans la plupart des situations réelles.