6

Comparer Modèles

6. Comparer modèles

La qualité d'une fonction est généralement évaluée à l'aide d'une fonction de perte.
Compte tenu d'un problème, d'un type de modèle et d'une fonction de perte choisis, la tâche de recherche du meilleur modèle est exprimée en un problème d'optimisation.
Ici, nous essayons de minimiser une fonction de coût par rapport aux paramètres du modèle (pour un modèle linéaire, les coefficients (ou poids) et le biais). Une fonction de coût est la moyenne des pertes calculées sur tous les ensembles de données d'apprentissage.
Fonctions de perte de régression :

1. Mean Squared Error (Erreur quadratique moyenne)/ perte quadratique (MSE)
2. Erreur absolue moyenne (MAE)
3. Smooth MAE
4. Log cosh Loss
5. sPerte quantile

Problème d'optimisation/ Minimisation des coûts : l'objectif est de trouver les paramètres du modèle (poids et biais) qui minimisent la fonction de coût.
L'optimisation est réalisée à l'aide d'algorithmes comme la descente de gradient, la descente de gradient stochastique, etc.

Partial derivative of the cost function w.r.t. j^{th} parameter

$$\frac{\delta}{\delta\theta_j} C_{MSE}(\theta) \qquad \frac{\delta}{\delta\theta_j} C_{LogLoss}(\theta)$$

Gradient vector of cost function: Partial derivative of the cost function all parameters

$$\nabla_\theta\, C_{MSE}(\theta) \qquad \nabla_\theta\, C_{LogLoss}(\theta)$$

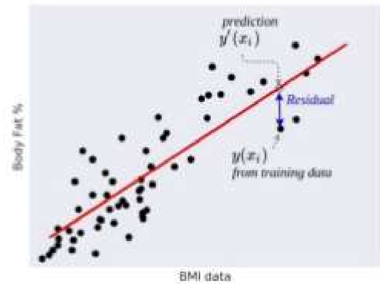
Gradient descent Step

$$\theta^{\text{next step}} = \theta - \eta\nabla_\theta\, C_{MSE}(\theta) \qquad \theta^{\text{next step}} = \theta - \eta\nabla_\theta\, C_{LogLoss}(\theta)$$

Learning rate/ Step size

REGRESSION

Ordinary Least Squares Regression



Residual/ error $= \text{training target value - predicted value} = y - y'$

Loss function $L(y, y') = \frac{1}{2}(y - y')^2$

Cost Function Mean Squared Error (**MSE**) $C(\theta) = \frac{1}{N}\Sigma_{i=1}^{N} L(y_i, y_i')$

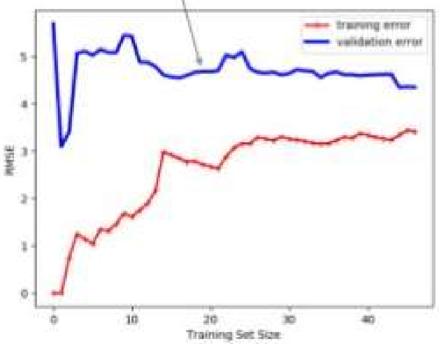
6. Comparer modèles

Mesure de la performance du modèle :

Regression performance measure & model evaluation	***Mathematical expression/ definition***
Mean Squared Error	$MSE(y,y') = \frac{1}{n_{samples}} \Sigma_{i=0}^{n_{samples}-1} (y_i - y'_i)^2$
Root Mean Squared Error	$RMSE(y,y') = \sqrt{MSE(y,y')}$
Mean Absolute Error	$MSE(y,y') = \frac{1}{n_{samples}} \Sigma_{i=0}^{n_{samples}-1} \lvert y_i - y'_i \rvert$
Explained Variance score Best possible score = 1.0, worsens with lower values	$exp.\ var(y,y') = 1 - \frac{Variance\ \{y-y'\}}{Variance\ \{y\}}$
R2 Score It provides an indication of goodness of fit and therefore a measure of how well unseen samples are likely to be predicted by the model.	$R^2(y,y') = 1 - \frac{\Sigma_{i=1}^{n} (y_i - y'_i)^2}{\Sigma_{i=1}^{n} (y - (\frac{1}{n}\Sigma_{i=1}^{n} (y_i)))^2}$
Cross Validation Evaluation *using Scikit learn's K fold cross validation* Estimates Model's generalization performance	Randomly split the training dataset into *K* distinct subsets called *folds* and it trains and evaluates the model *K* times, picking a different fold for evaluation every time and training on the other *K-1* folds.
Learning Curves	Plots a model's performance on the training dataset and validation dataset as a function of training set size.

Learning curves for our regression model

How model generalizes with increasing training dataset size



Mean Square Error (MSE) L2 est la fonction de perte de régression la plus couramment utilisée. MSE est la somme des distances au carré entre notre variable cible et les valeurs prédites.

Mean Absolute Error (MAE) L1 est une autre fonction de perte utilisée pour les modèles de régression. MAE est la somme des différences absolues entre nos variables cibles et prédites. Elle mesure donc la magnitude moyenne des erreurs dans un ensemble de prédictions, sans tenir compte de leurs directions. Plus robustes aux outliers. PB son gradient est le même partout, ce qui signifie que le gradient sera important même pour de petites valeurs de perte. Ce n'est pas bon pour l'apprentissage. Pas utiliser pour les réseaux de neurones. Si les valeurs aberrantes représentent des anomalies importantes pour l'entreprise et doivent être détectées, alors nous devrions utiliser MSE. En revanche, si nous pensons que les valeurs aberrantes représentent simplement des données corrompues, nous devrions choisir MAE comme perte.

6. Comparer modèles

Mesure de la performance du modèle :

Choix de la métrique:

1. **MAE : Mean Average Error** — Mesure de l'erreur moyenne réalisée
$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

2. **RMSE : Root Mean Squared Error** — Mesure de l'erreur pénalisant les grandes erreurs
$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

3. **Root Mean Squared Logarithmic Error** — Pénalisation des valeurs sous-estimant la valeur cible
$$\epsilon = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_i + 1) - \log(a_i + 1))^2}$$

4. **Autres mesures** — Mean Squared Error, Weigthed Mean Average Error, ...
$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$$

Le R² ajusté est une version modifiée du R², et il est ajusté pour le nombre de variables indépendantes dans le modèle, et il sera toujours inférieur ou égal au R². Dans la formule ci-dessous, n est le nombre d'observations dans les données et k est le nombre de variables indépendantes dans les données.

$$R^2_{adj} = 1 - \left[\frac{(1 - R^2)(n-1)}{n-k-1}\right]$$

MAE: l'erreur absolue moyenne représente la moyenne de la différence absolue entre les valeurs réelles et prédites dans l'ensemble de données. Elle mesure la moyenne des résidus dans l'ensemble de données.

MSE : l'erreur quadratique moyenne représente la moyenne de la différence au carré entre les valeurs originales et prédites dans l'ensemble de données. Elle mesure la variance des résidus.

RMSE : l'erreur quadratique moyenne est la racine carrée de l'erreur quadratique moyenne. Elle mesure l'écart-type des résidus.

R2 : Le coefficient de détermination ou R-carré représente la proportion de la variance de la variable dépendante qui est expliquée par le modèle de régression linéaire. Il s'agit d'un score sans échelle, c'est-à-dire qu'indépendamment du fait que les valeurs soient petites ou grandes, la valeur du carré R sera inférieure à un.

6. Comparer modèles

Différences entre ces mesures d'évaluation :

- MSE et RMSE pénalisent les erreurs de prédiction importantes par rapport à la MAE. Cependant, RMSE est plus largement utilisée que MSE pour évaluer la performance du modèle de régression avec d'autres modèles aléatoires car elle a les mêmes unités que la variable dépendante (axe Y).

- La MSE est une fonction différentiable qui facilite les opérations mathématiques par rapport à une fonction non différentiable comme la MAE. Par conséquent, dans de nombreux modèles, lRMSE est utilisée comme métrique par défaut pour calculer la fonction de perte, bien qu'elle soit plus difficile à interpréter que la MAE.

- MAE est plus robuste aux données comportant des valeurs aberrantes.

- Une valeur plus faible de MAE, MSE et RMSE implique une plus grande précision d'un modèle de régression. Cependant, une valeur plus élevée du R carré est considérée comme souhaitable.

- Le R au carré et le R au carré ajusté sont utilisés pour expliquer dans quelle mesure les variables indépendantes du modèle de régression linéaire expliquent la variabilité de la variable dépendante. La valeur du R carré augmente toujours avec l'ajout des variables indépendantes, ce qui pourrait conduire à l'ajout de variables redondantes dans notre modèle. Cependant, le R au carré ajusté résout ce problème.

- Le R au carré ajusté prend en compte le nombre de variables prédictives, et il est utilisé pour déterminer le nombre de variables indépendantes dans notre modèle. La valeur du R au carré ajusté diminue si l'augmentation du R au carré par la variable supplémentaire n'est pas assez significative.

- Pour comparer la précision de différents modèles de régression linéaire, la RMSE est un meilleur choix que le R au carré.

- Par conséquent, si l'on compare la précision de prédiction entre différents modèles de régression linéaire (LR), la RMSE est une meilleure option car elle est simple à calculer et différentiable. Toutefois, si votre ensemble de données comporte des valeurs aberrantes, choisissez MAE plutôt que RMSE.

- En outre, le nombre de variables prédicteurs dans un modèle de régression linéaire est déterminé par le R au carré ajusté, et choisissez RMSE plutôt que R au carré ajusté si vous vous souciez d'évaluer la précision de la prédiction parmi différents modèles LR.

Coefficient of Determination (R Square)

$$R^2 = \frac{SSR}{SST}$$

$$SSR = \sum_i (\hat{y}_i - \bar{y})^2$$

$$SST = \sum_i (y_i - \bar{y})^2$$

Where,

- SSR is Sum of Squared Regression also known as variation explained by the model
- SST is Total variation in the data also known as sum of squared total
- y_i is the y value for observation i
- y_bar is the mean of y value
- y_bar_hat is predicted value of y for observation i