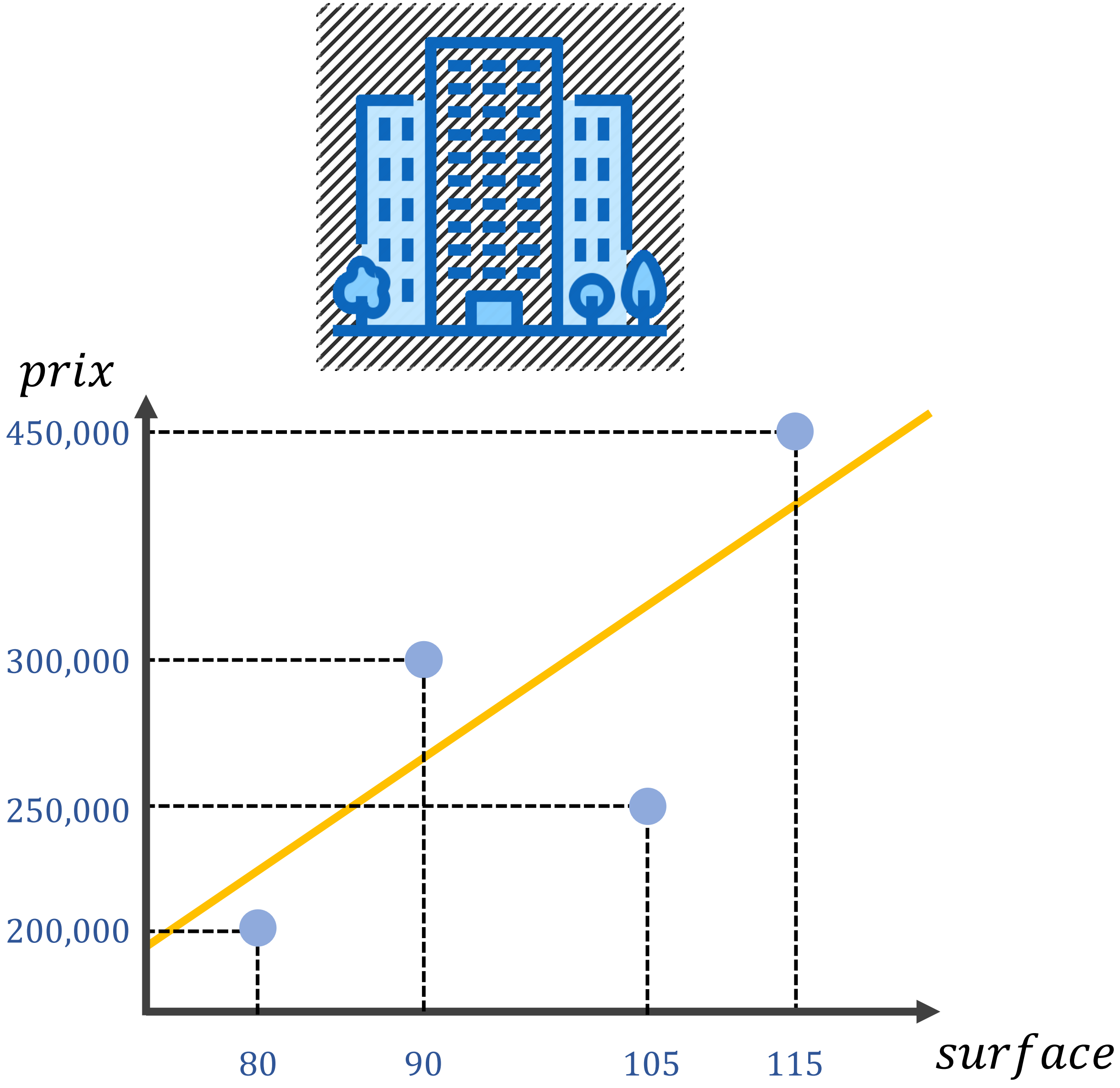
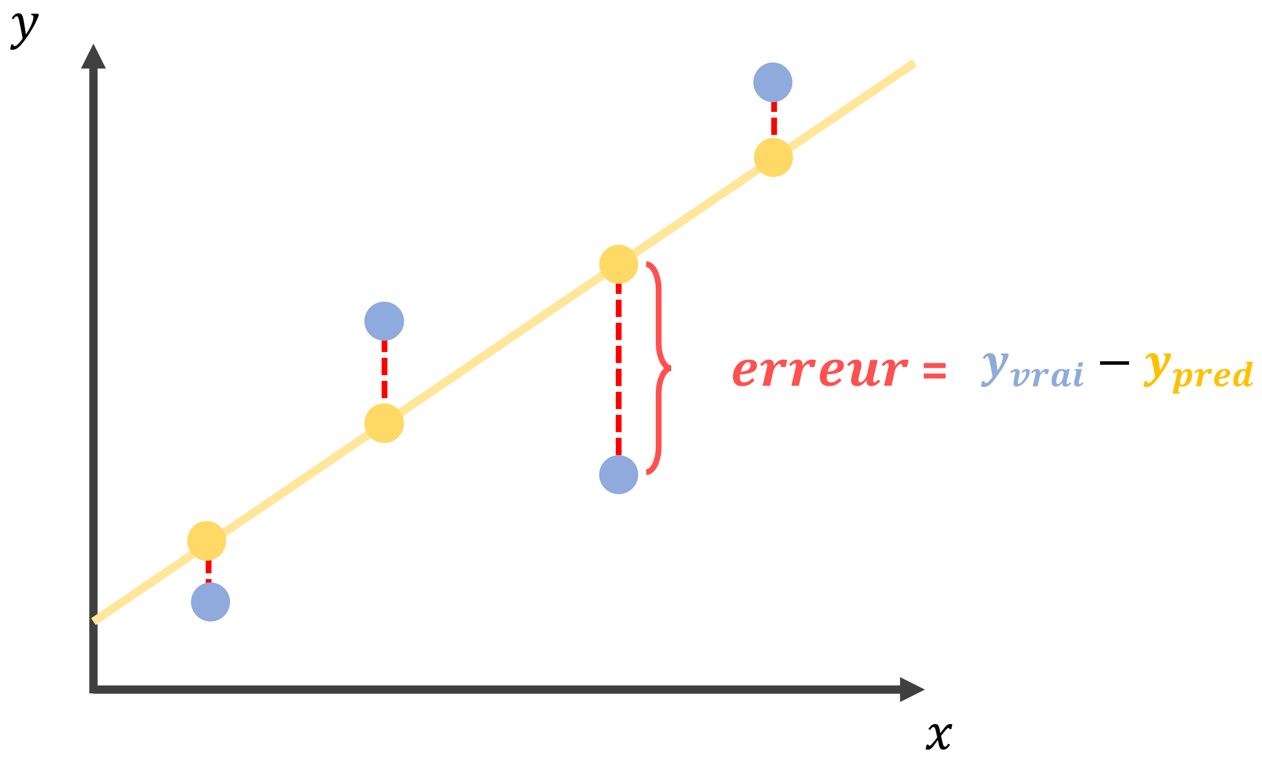
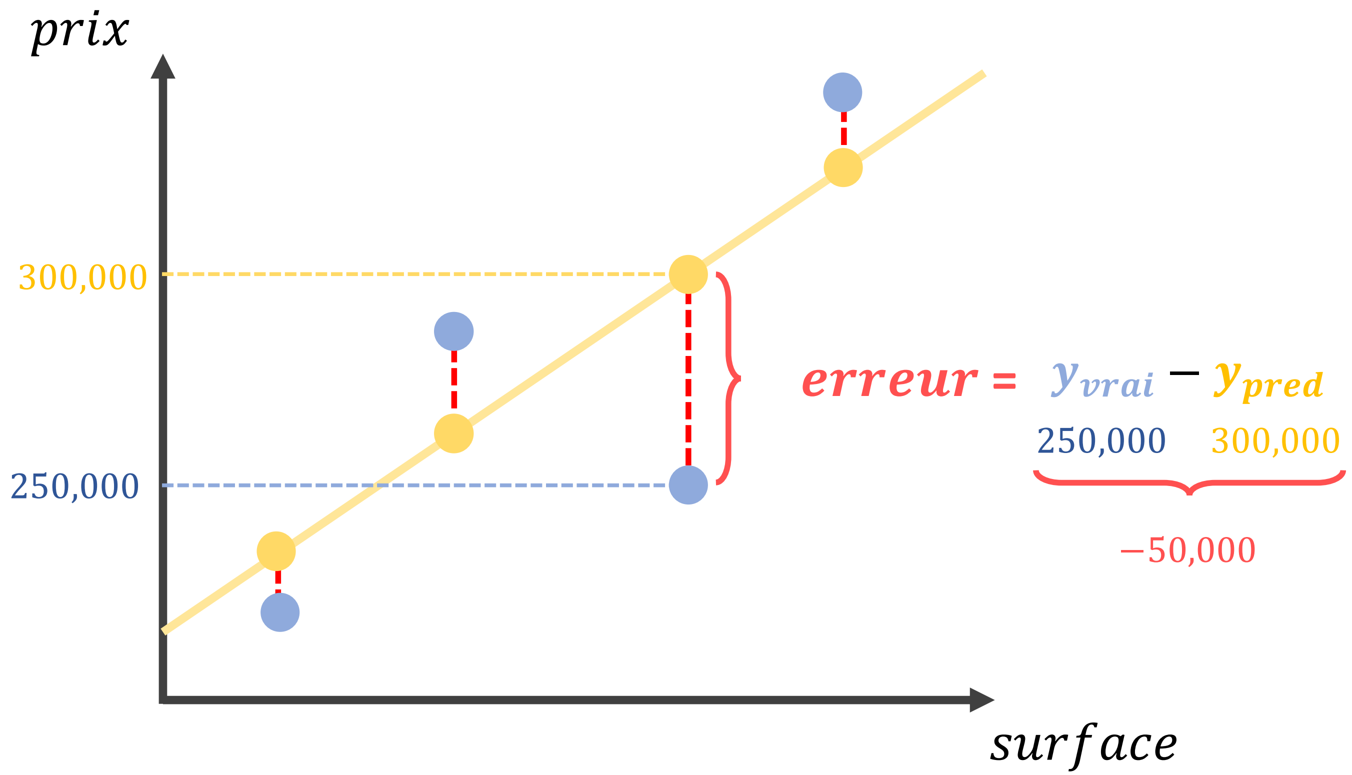
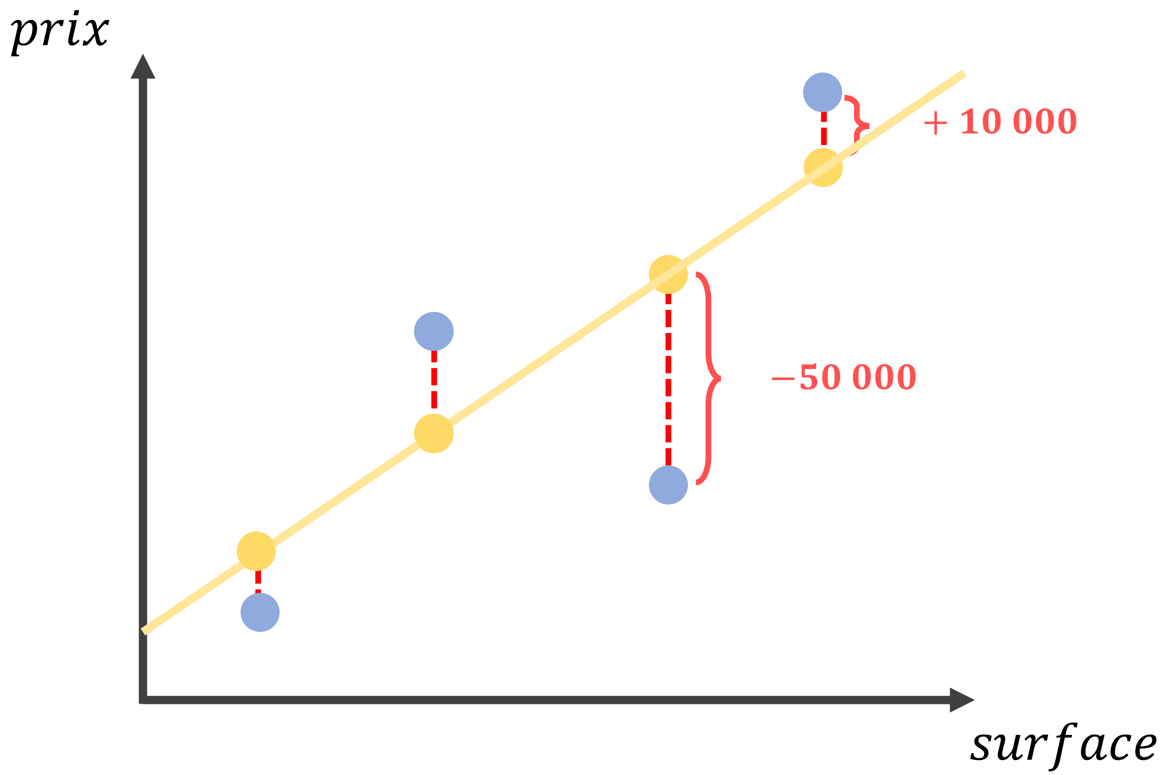

prix
450,000
300,000
250,000
200,000
80
90
105
115
surface

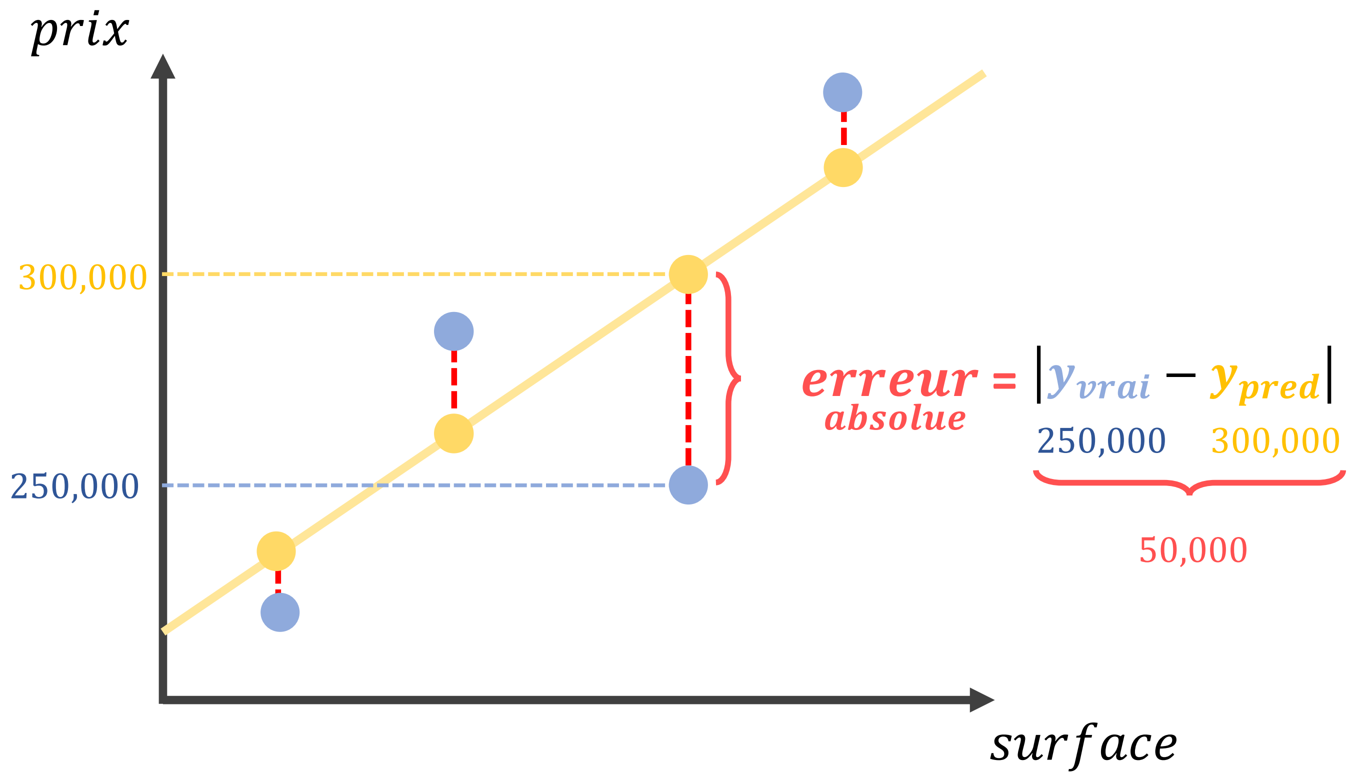


y
x
erreur = y_vrai − y_pred

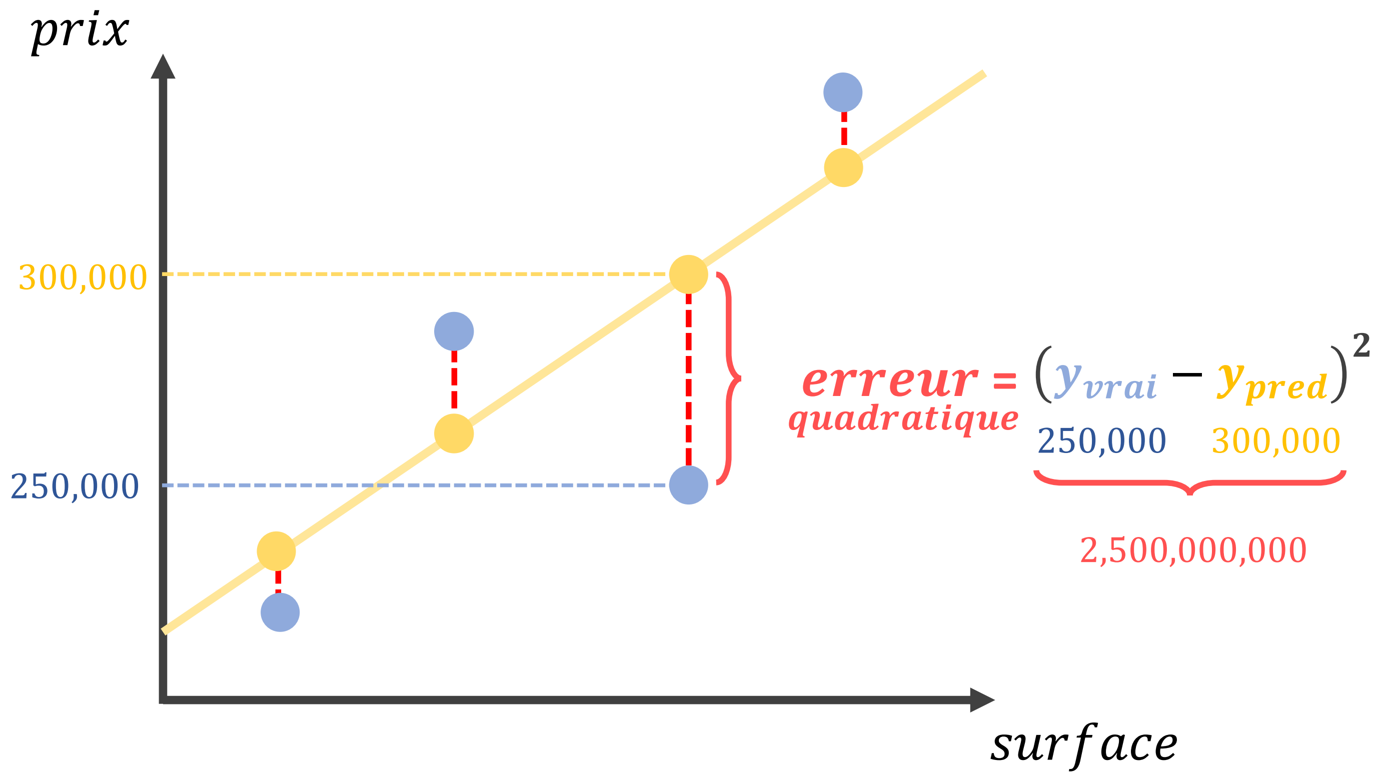


prix
300,000
250,000
erreur = $y_{vrai} - y_{pred}$
250,000
300,000
−50,000
surface


prix
+ 10 000
−50 000
surface

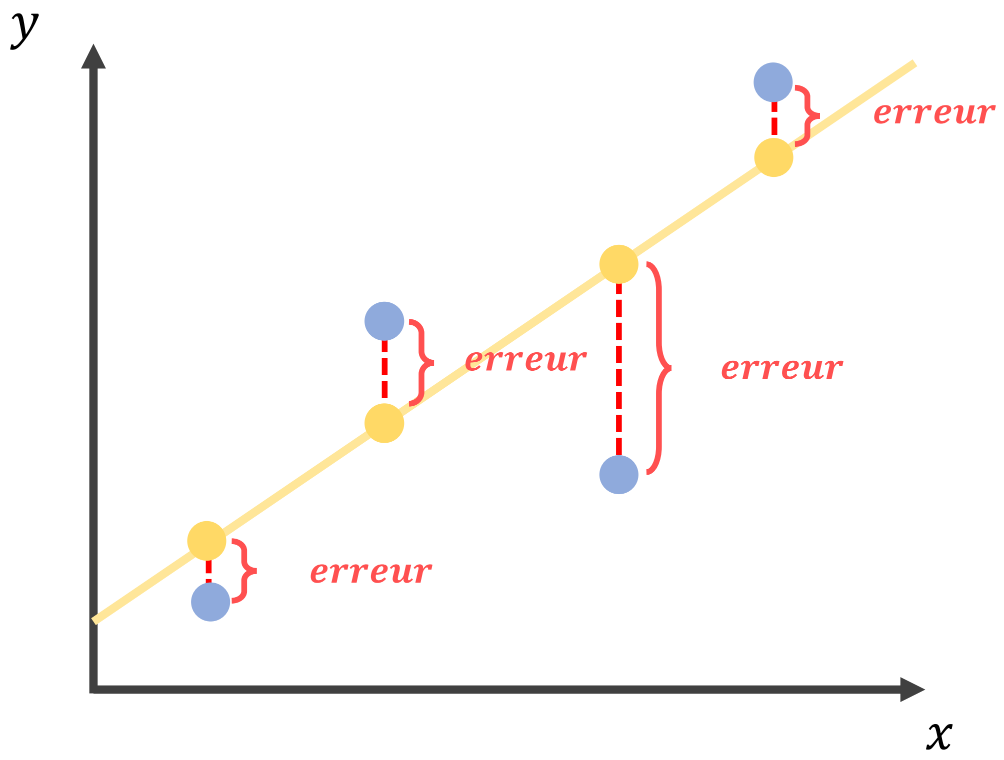


prix
300,000
250,000
surface
erreur absolue = |y_vrai − y_pred|
250,000
300,000
50,000


prix
surface
300,000
250,000
erreur quadratique = $(y_{vrai} - y_{pred})^2$
250,000
300,000
2,500,000,000

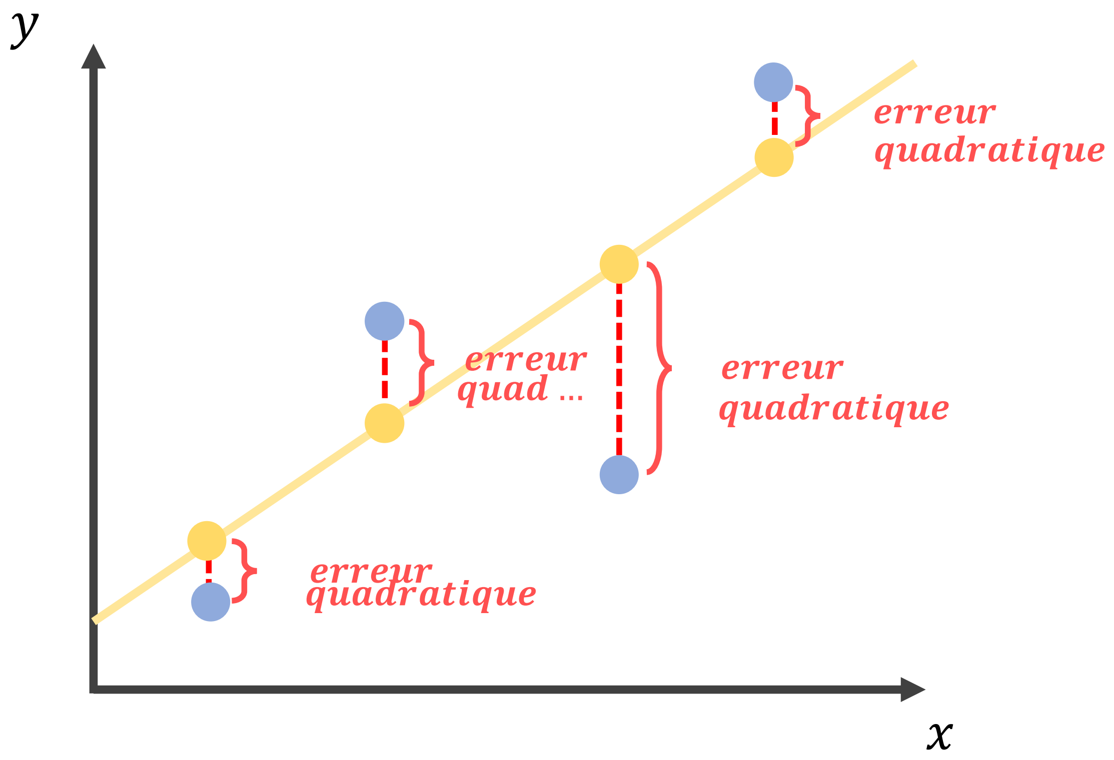
*la moyenne des erreurs*

$$\frac{1}{m}\sum \textit{erreur}$$



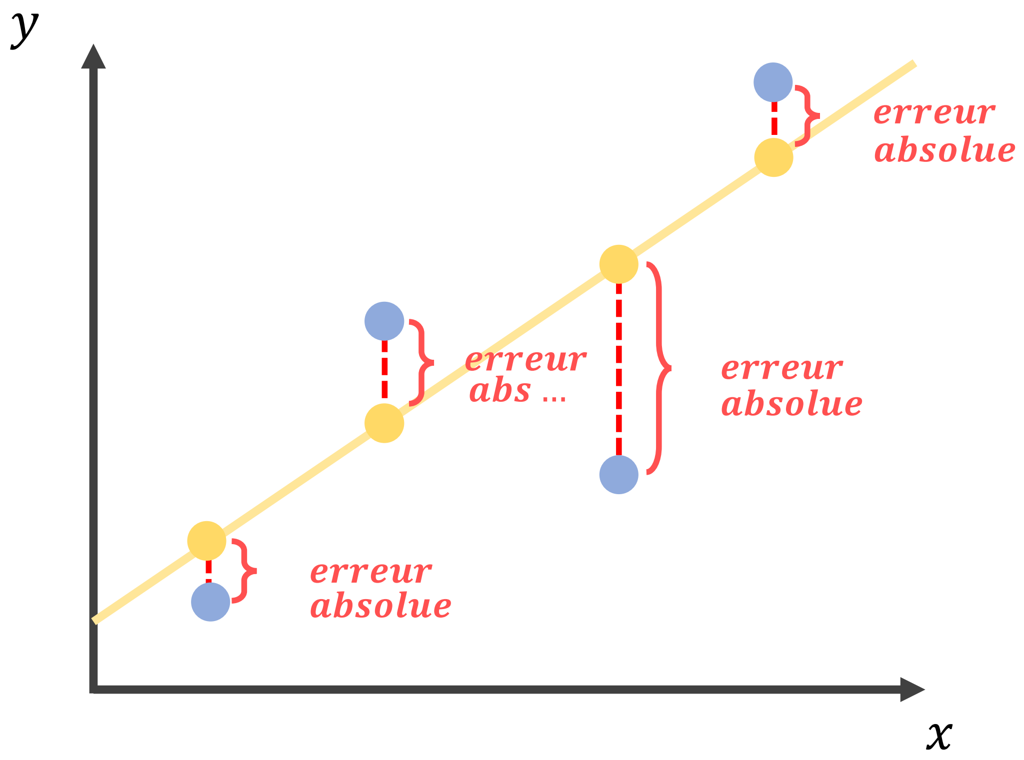
## *Erreur quadratique moyenne*

$$MSE = \frac{1}{m}\sum\left(y_{vrai} - y_{pred}\right)^2$$



## *Erreur absolue moyenne*

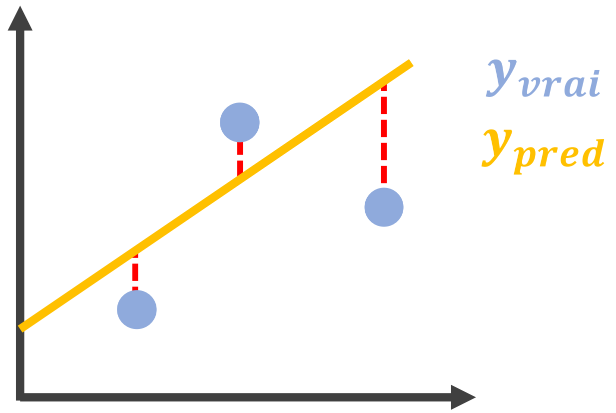
$$MAE = \frac{1}{m}\sum \left| y_{vrai} - y_{pred} \right|$$



# MAE
# MSE
# RMSE

$$MAE = \frac{1}{m}\sum\left|y_{vrai} - y_{pred}\right|$$

$$MSE = \frac{1}{m}\sum\left(y_{vrai} - y_{pred}\right)^2$$



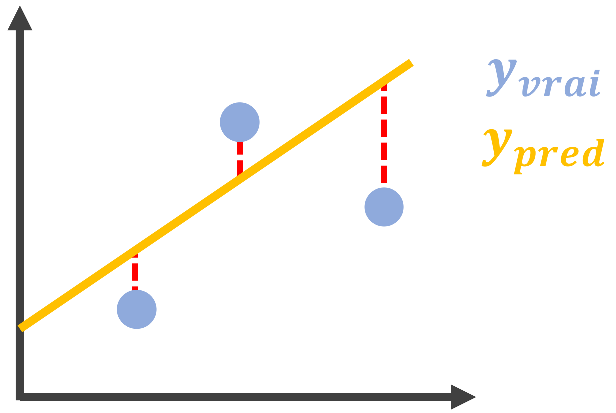
# MAE
# MSE
# RMSE

$$MAE = \frac{1}{m}\sum\left|y_{vrai} - y_{pred}\right|$$

$$|1 - 1| = 0$$

$$MSE = \frac{1}{m}\sum\left(y_{vrai} - y_{pred}\right)^2$$

$$(1 - 1)^2 = 0$$



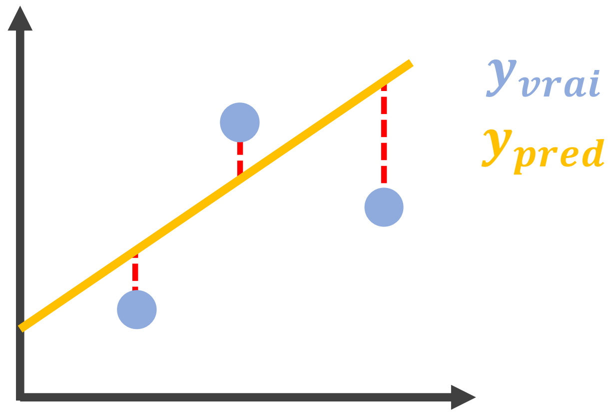
**MAE**

**MSE**

**RMSE**

$$MAE = \frac{1}{m}\sum\left|y_{vrai} - y_{pred}\right| \qquad |1-2| = 1$$

$$MSE = \frac{1}{m}\sum\left(y_{vrai} - y_{pred}\right)^2 \qquad (1-2)^2 = 1$$



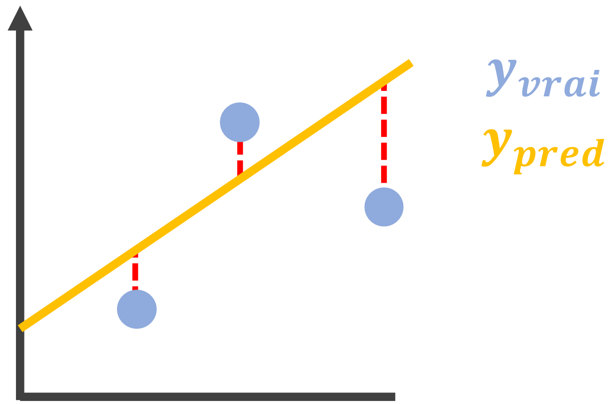
# MAE
# MSE
# RMSE

$$MAE = \frac{1}{m}\sum\left|y_{vrai} - y_{pred}\right|$$

$$|1 - 3| = 2$$

$$MSE = \frac{1}{m}\sum\left(y_{vrai} - y_{pred}\right)^2$$

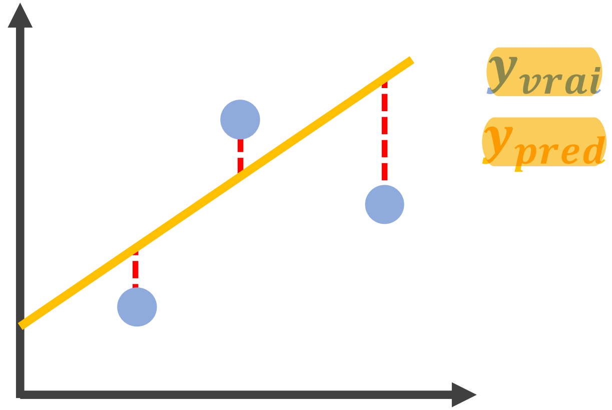
$$(1 - 3)^2 = 4$$



# MAE
# MSE
# RMSE

$$MAE = \frac{1}{m}\sum \left| y_{vrai} - y_{pred} \right| \qquad |1-4| = 3$$

$$MSE = \frac{1}{m}\sum \left( y_{vrai} - y_{pred} \right)^2 \qquad (1-4)^2 = 9$$



$$\boxed{\sum (erreurs)^2 \neq \left(\sum erreurs\right)^2}$$

$$\boxed{MSE \neq (MAE)^2}$$

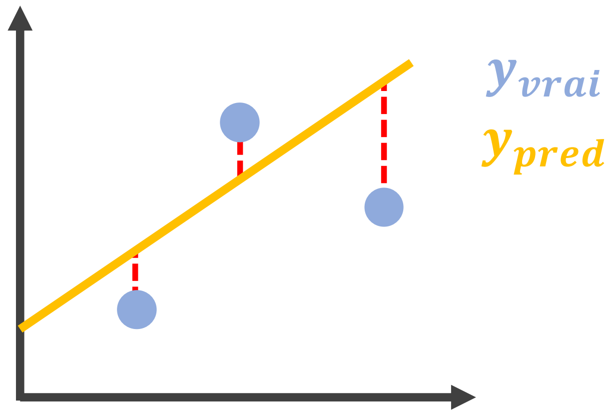
# MAE
# MSE
# RMSE

$$MAE = \frac{1}{m}\sum \left|y_{vrai} - y_{pred}\right|$$

$$\frac{4+0}{2} = 2$$

$$MSE = \frac{1}{m}\sum \left(y_{vrai} - y_{pred}\right)^2$$

$$\frac{4^2+0^2}{2} = 8$$



# MAE
# MSE
# RMSE

$$MAE = \frac{1}{m}\sum\left|y_{vrai} - y_{pred}\right|$$

$$\frac{4+0}{2} = 2$$

$$MSE = \frac{1}{m}\sum\left(y_{vrai} - y_{pred}\right)^2$$

$$\frac{4^2+0^2}{2} = 8$$

$$RMSE = \sqrt{\frac{1}{m}\sum\left(y_{vrai} - y_{pred}\right)^2}$$

$$\sqrt{\frac{4^2+0^2}{2}} = 2.8$$

La RMSE retourne l'erreur à **son échelle initiale,** ce qui est plus compréhensible.

***Mean Absolute Error***

$$\frac{4+0}{2}=\mathbf{2}$$

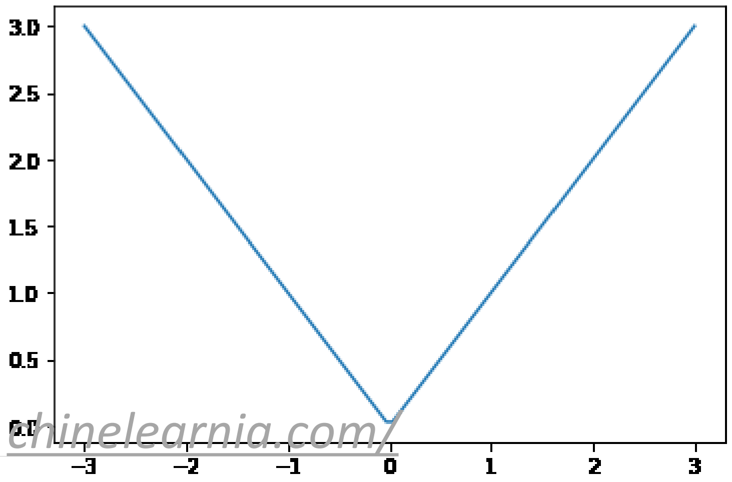
***Root Mean Squared Error***

$$\sqrt{\frac{4^2+0^2}{2}}=\mathbf{2.8}$$

# Quand utiliser la **MSE** plutôt que la **MAE** ?

***Mean Absolute Error***

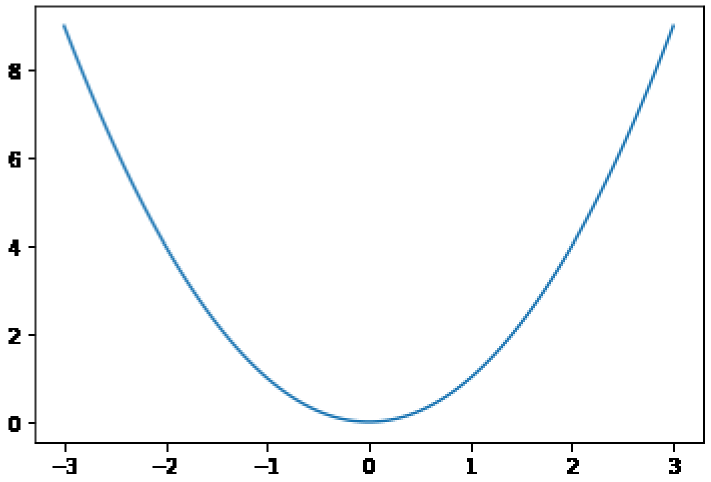
l'importance d'une erreur est **linéaire** avec son amplitude






***Root Mean Squared Error***

l'importance d'une erreur est **exponentielle** avec son amplitude



# Estimateur de distance de freinage


| | $Erreur\ 1$ | $Erreur\ 2$ | $\boldsymbol{MAE}$ | $\boldsymbol{MSE}$ |
|---|---|---|---|---|
| **Modèle A** | 10 | 0 | **5** | |
| ~~Modèle B~~ | 6 | 5 | 5.5 | |

# Estimateur de distance de freinage


| | *Erreur 2* | ***MAE*** | ***MSE*** |
|---|---|---|---|
| **Modèle A** | 0 | **5** | |
| Modèle B | 5 | 5.5 | |

Est-ce vraiment une bonne chose de choisir le modèle A ? Il peut être **très dangereux**….

# Estimateur de distance de freinage


| | $Erreur\ 1$ | $Erreur\ 2$ | $\boldsymbol{MAE}$ | $\boldsymbol{MSE}$ |
|---|---|---|---|---|
| ~~Modèle A~~ | 10 | 0 | 5 | 7 |
| **Modèle B** | 6 | 5 | 5.5 | **5.52** |

# Estimateur de distance de freinage


Une erreur de **10 mètres**, c'est **100 fois plus grave** qu'une erreur de **1 mètre.**

| | $Erreur\ 2$ | $\boldsymbol{MAE}$ | $\boldsymbol{MSE}$ |
|---|---|---|---|
| ~~Modèle A~~ | 0 | 5 | 7 |
| **Modèle B** | 5 | 5.5 | **5.52** |

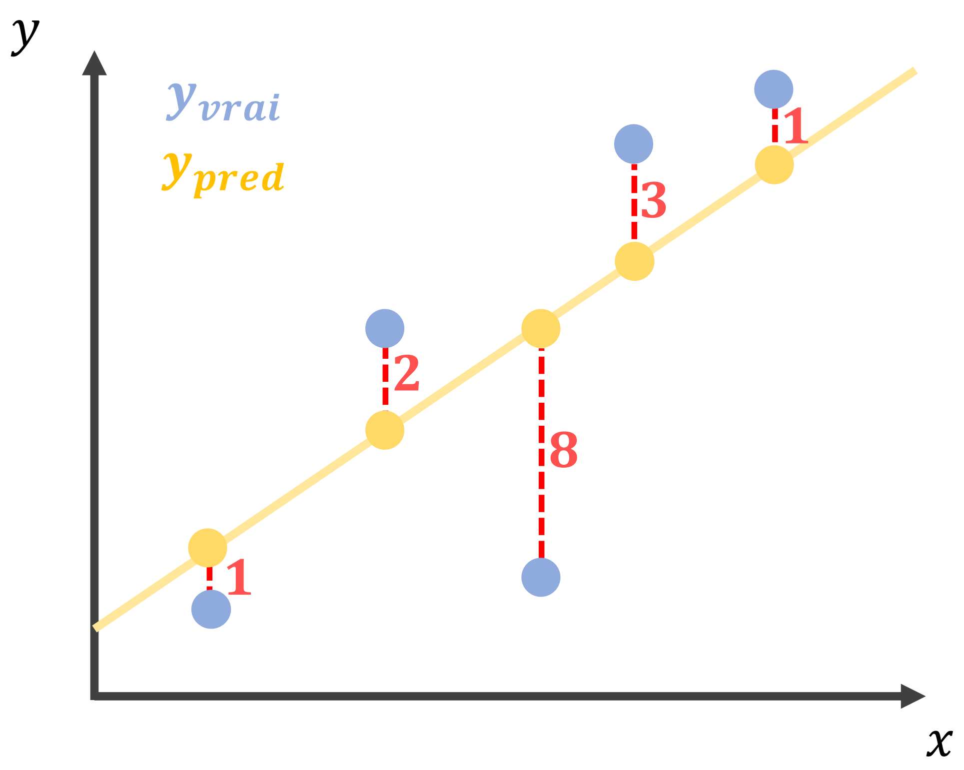
la MSE **pénalise beaucoup plus** les **grandes erreurs** que la MAE.

# Quand utiliser la **MSE** plutôt que la **MAE** ?

$\boldsymbol{MSE}$ → vous accordez une **grande importance** aux **grandes erreurs.**

$\boldsymbol{MAE}$ → l'importance d'une erreur est **linéaire** avec son amplitude. Si le Dataset contient des valeurs **aberrantes** (*outliers*).

# Median Absolute Error




$$erreurs = \{1, 1, 2, 3, 8\}$$

*median*

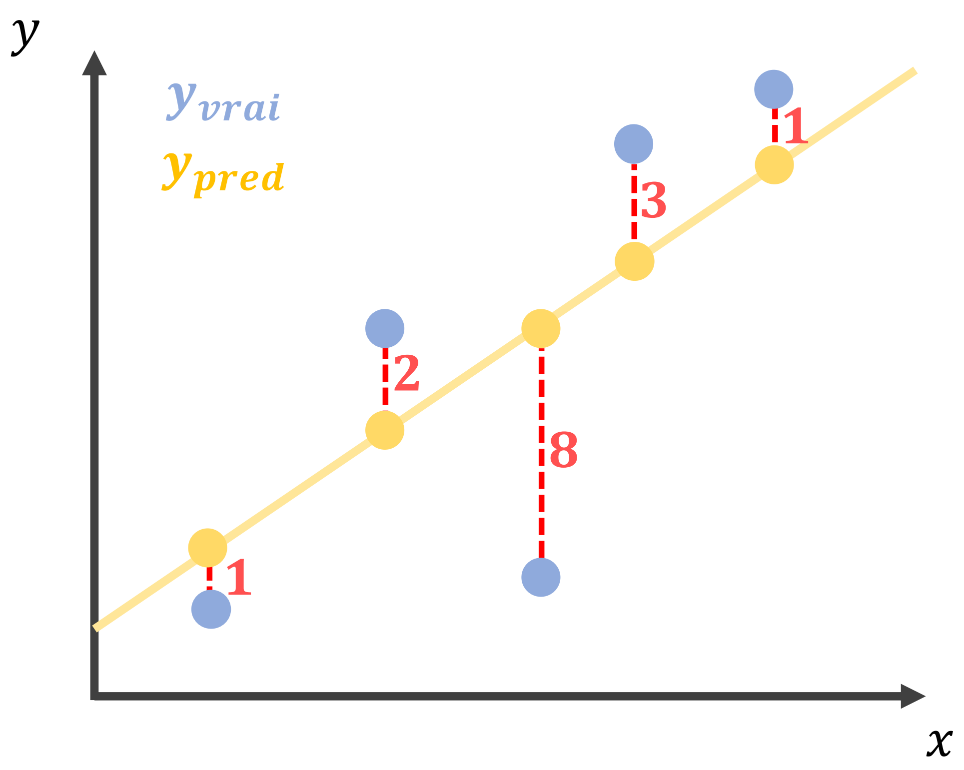
$$\boxed{Median\ Absolute\ Error = 2}$$

Il existe aussi la ***Median Absolute Error***.

**Très peu sensible** aux grandes erreurs.

$$MAE = median\,\{|y_{vrai} - y_{pred}|\}$$

**Attention** quand vous l'utilisez….




$$erreurs = \{\, 1, 1, \mathbf{2}, 3, 8 \,\}$$

*median*

$$\boxed{Median\ Absolute\ Error = 2}$$

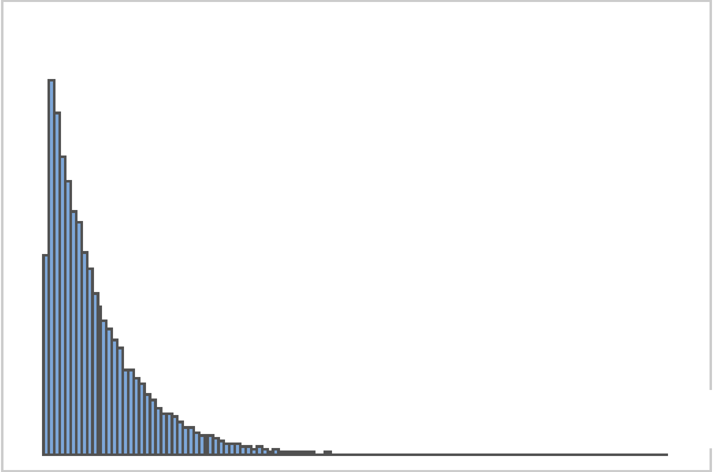
# Au Final…

MAE, MSE, RMSE, Médiane…

Il semble que chaque mesure est des **avantages** et des **inconvénients**… Du coup, laquelle choisir ?

→ **Utilisez les toutes !**

Vous récolterez ainsi beaucoup plus **d’information !**



# Coefficient de Determination R2

$$R2 = 1 - \frac{\overbrace{\boxed{\sum\left(y_{vrai} - y_{pred}\right)^2}}^{erreur}}{\underbrace{\boxed{\sum\left(y_{vrai} - \overline{y_{vrai}}\right)^2}}_{variance}}$$

*Évalue la **performance** du modèle **par rapport** au **niveau de variation** présent dans les données.*

# Coefficient de détermination R2

$$\boldsymbol{R2 = 1 - \frac{\overbrace{\boxed{\sum\left(y_{vrai} - y_{pred}\right)^2}}^{erreur\ quadratique}}{\underbrace{\boxed{\sum\left(y_{vrai} - \overline{y_{vrai}}\right)^2}}_{variance}}}$$

moyenne des $y_{vrai}$

*Évalue la **performance** du modèle <u>**par rapport**</u> au **niveau de variation** présent dans les données.*

# Coefficient de détermination R2

$$\underset{\mathbf{1}}{\boxed{\boldsymbol{R2}}} = \mathbf{1} - \underset{\mathbf{0}}{\boxed{\frac{\boldsymbol{erreurs}}{\boldsymbol{variance}}}}$$

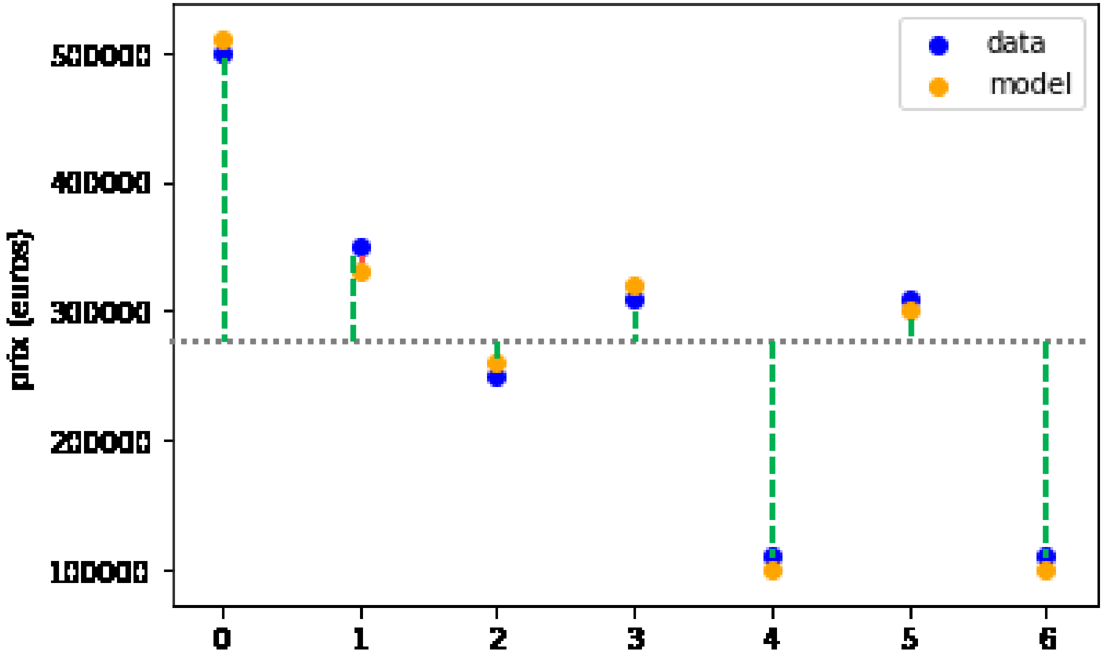
$$\boldsymbol{erreurs} \ll \boldsymbol{variance}$$

# Coefficient de détermination R2

$$R2 = 1 - \frac{\overbrace{\sum(y_{vrai} - y_{pred})^2}^{1000}}{\underbrace{\sum(y_{vrai} - \overline{y_{vrai}})^2}_{100000}}$$

$$R2 = 1 - 0.01$$

$$\boxed{R2 = 0.99}$$

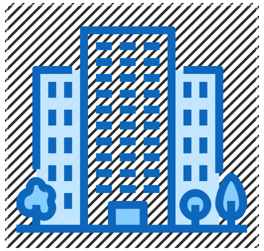


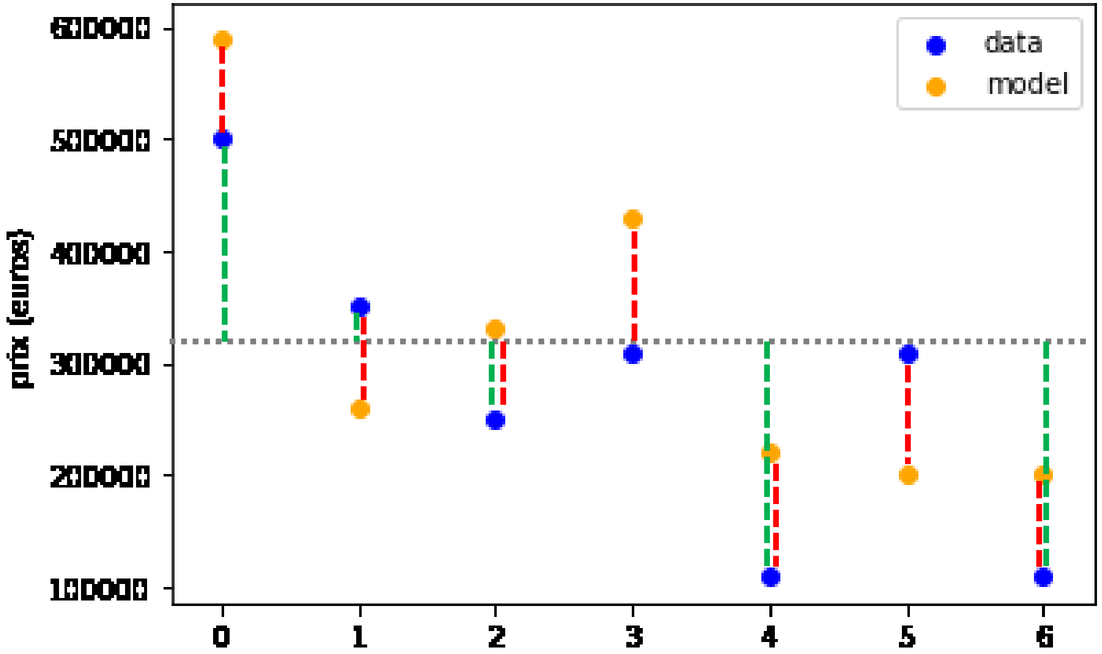
# Coefficient de détermination R2

$$R2 = 1 - \frac{\overbrace{\sum(y_{vrai} - y_{pred})^2}^{100000}}{\underbrace{\sum(y_{vrai} - \overline{y_{vrai}})^2}_{100000}}$$

$$R2 = 1 - 1$$

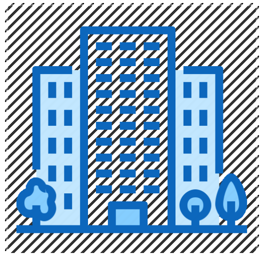
$$\boxed{R2 = 0}$$

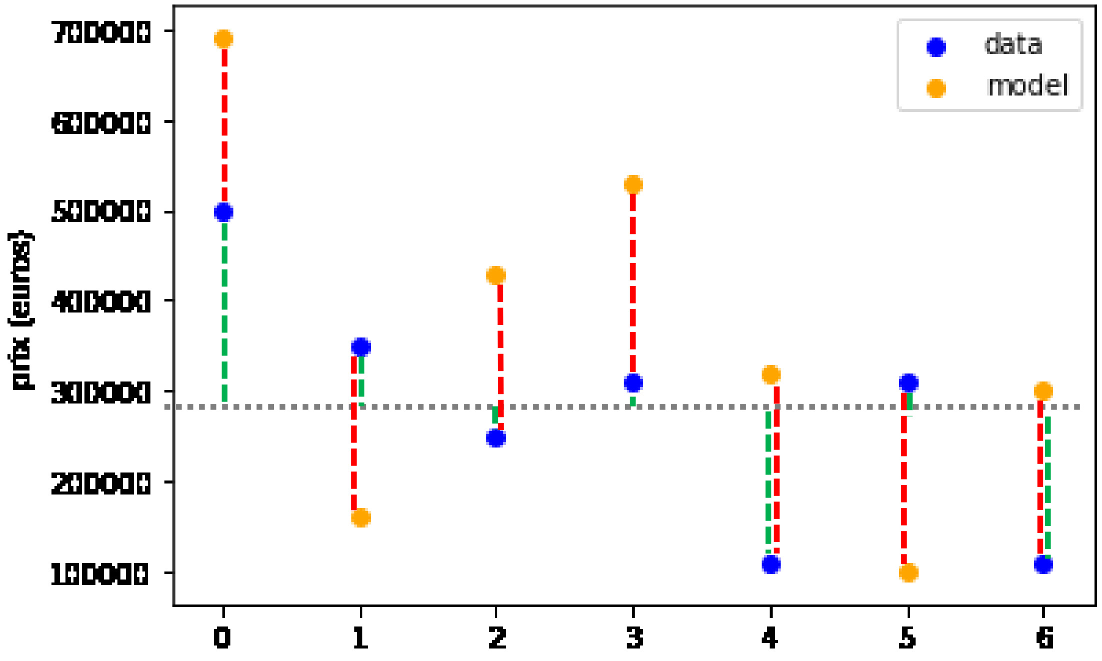




# Coefficient de détermination R2

$$R2 = 1 - \frac{\overbrace{\sum(y_{vrai} - y_{pred})^2}^{200000}}{\underbrace{\sum(y_{vrai} - \overline{y_{vrai}})^2}_{100000}}$$






$$R2 = 1 - 2$$

$$\boxed{R2 = -1}$$

# Quelle formule peut sauver ce piéton ?

| MAE | MSE |
| --- | --- |
| $\frac{1}{m}\sum\left\|y_{vrai} - y_{pred}\right\|$ | $\frac{1}{m}\sum\left(y_{vrai} - y_{pred}\right)^2$ |