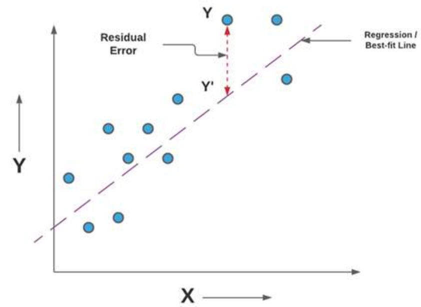
Performance indices

Regression

Jean Christophe Meunier
January the 4th 2021, Becode AI/data science bootcamp

Error-based indicators

Characteristics

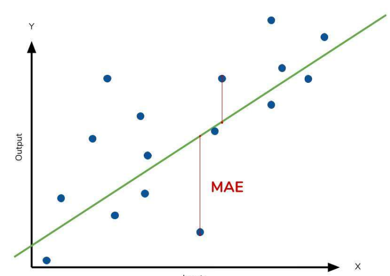


- Distance observed (y) -- predicted (y' or ŷ)
  - The lower the best
- Not readily interpretable, but useful :
  - For comparing nested models (error is data set dependent)
  - In conjunction with other indicators
- Many indicators :
  - E.g. MAE, MSE, RMSE, MPE, MPA,..
  - Each of them covering different aspects : amount of error, bias, specific outliers,…

Mean absolute error (MAE)

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

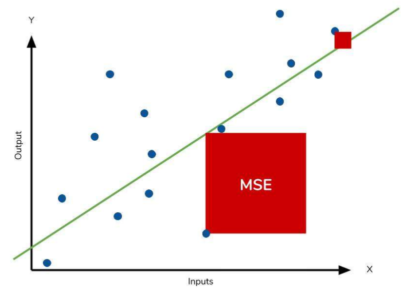
- Absolute distance :
  - + and – distance do not cancel each other out
- Typical magnitude of the residuals
  - Same metric as variable of interest



Root Mean Squared Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

- Distance ‘manipulation’ :
  1. Squared :
     - - and + do not cancel each other out
     - larger error are exponentially weighted
  2. Square root :
     - Turn back in the same metric as variable of interest



MAE vs. RMSE

- MAE
  - Easier to interpret (distance is not ‘manipulated’)
  - Errors contribute linearly to the overall error
- RMSE
  - More useful when large errors are particularly undesirable (outliers)

CASE 1: Evenly distributed errors

ID	Error	|Error|	Error^2
1	2	2	4
2	2	2	4
3	2	2	4
4	2	2	4
5	2	2	4
6	2	2	4
7	2	2	4
8	2	2	4
9	2	2	4
10	2	2	4

MAE	RMSE
2.000	2.000

CASE 2: Small variance in errors

ID	Error	|Error|	Error^2
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	1	1	1
6	3	3	9
7	3	3	9
8	3	3	9
9	3	3	9
10	3	3	9

MAE	RMSE
2.000	2.236

CASE 3: Large error outlier

ID	Error	|Error|	Error^2
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	20	20	400

MAE	RMSE
2.000	6.325

MAE vs. RMSE

- Both provide information on amount of error but <u>not</u> on bias

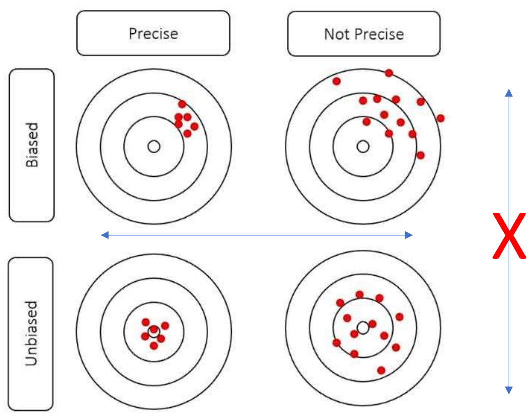


Figure 2.1: Precision & Bias

with **|abs|** and **()²**

+ and – distances are no longer identifiable

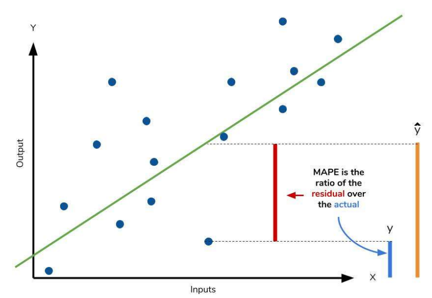
Mean Absolute Percentage of Error (MAPE)

- Average proportion of error
  - Ex: $\hat{y}$ = 11 vs. y = 10 indicates a 10% error
- Independent of the metric
  - Ex : 1 vs. 10 same as 100 vs. 1000
- Absolute value
  - - and + distances do not cancel each other out

Multiplying by 100% converts to percentage

$$MAPE = \frac{\boxed{100\%}}{n} \sum \left| \frac{\overbrace{y-\hat{y}}^{\text{The residual}}}{\boxed{y}} \right|$$

Each residual is scaled against the actual value



Mean Absolute Percentage of Error (MAPE)

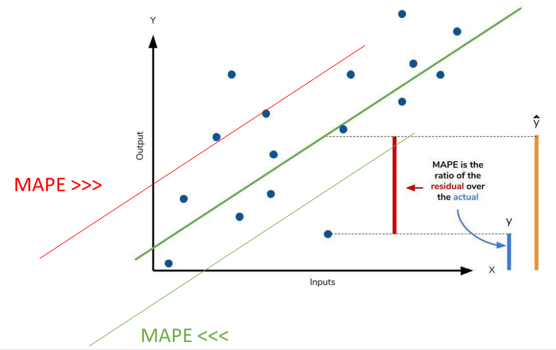
- Average proportion of error
  - Ex: $\hat{y}$ = 11 vs. y = 10 indicates a 10% error
- Independent of the metric
  - Ex : 1 vs. 10 same as 100 vs. 1000
- Absolute value
  - \- and + distances do not cancel each other out

Multiplying by 100% converts to percentage

The residual

$$MAPE = \frac{\boxed{100\%}}{n} \sum \left| \frac{\overbrace{y - \hat{y}}}{\boxed{y}} \right|$$

Each residual is scaled against the actual value



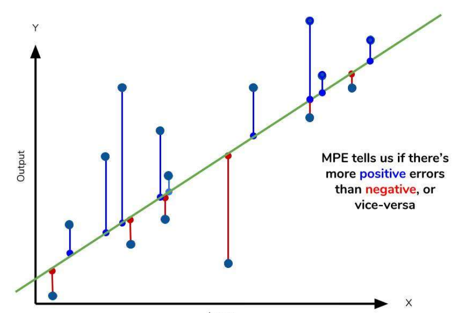
Weakness of MAPE: predictions that are systematically either higher or lower are not equally weighted

$\hat{y}$ is smaller than the actual value	$\hat{y}$ is greater than the actual value
n = 1 $\hat{y}$ = 10 y = 20	n = 1 $\hat{y}$ = 20 y = 10
MAPE = 50%	MAPE = 100%

Mean Percentage of Error (MPE)

- Same as MAPE but without |abs|
  - \+ and – canceled out
- No information on performance
- Identify bias
  - underestimation (more negative error)
  - overestimation (more positive error)

$$MPE = \frac{100\%}{n} \sum \left(\frac{y - \hat{y}}{y} \right)$$



Mean Percentage of Error (MPE)

- Provide information on bias but <u>not</u> on amount of error

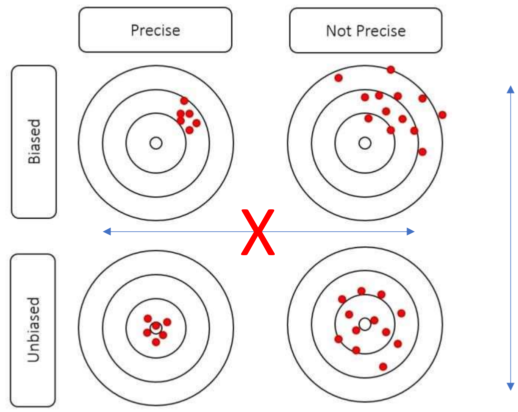


Figure 2.1: Precision & Bias

without **|abs|**

+ and – distances are now identifiable

Alternative indicators

- Median absolute error

$$MedAE(y,\hat{y}) = median(|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|)$$

  - Not affected by outliers

- Max error
  - worst case error between the predicted value and the true value

$$\text{Max Error}(y,\hat{y}) = max(|y_i - \hat{y}_i|)$$

Sklearn

```
Entrée [40]: from sklearn.metrics import mean_absolute_error
             from sklearn.metrics import mean_squared_error
             from sklearn.metrics import mean_absolute_percentage_error
             from sklearn.metrics import median_absolute_error
             from sklearn.metrics import max_error

Entrée [41]: mean_absolute_error(y_ess1, y_pred1)

   Out[41]: 0.7406040153701986

Entrée [42]: mean_squared_error(y_ess1, y_pred1)

   Out[42]: 0.8623720997604472

Entrée [43]: mean_absolute_percentage_error(y_ess1, y_pred1)

   Out[43]: 0.5803582139093125
```

New in sklearn 0,24

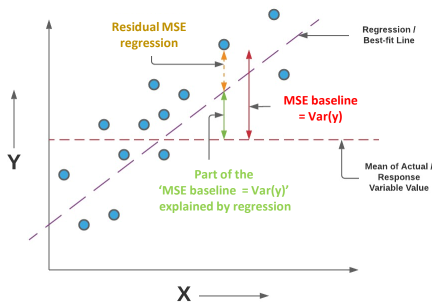
```
Entrée [44]: median_absolute_error(y_ess1, y_pred1)

   Out[44]: 0.6155660119853823

Entrée [45]: max_error(y_ess1, y_pred1)

   Out[45]: 2.742975648344551
```

Rem:
Mean percentage error (MPE) not available in sklearn

Scores

Based on explained variance

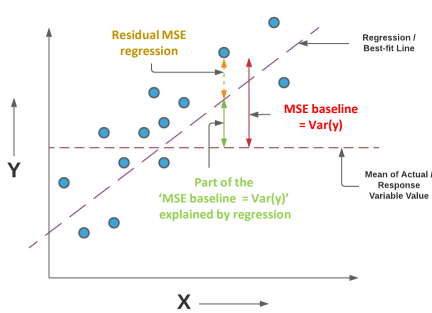
Characteristics



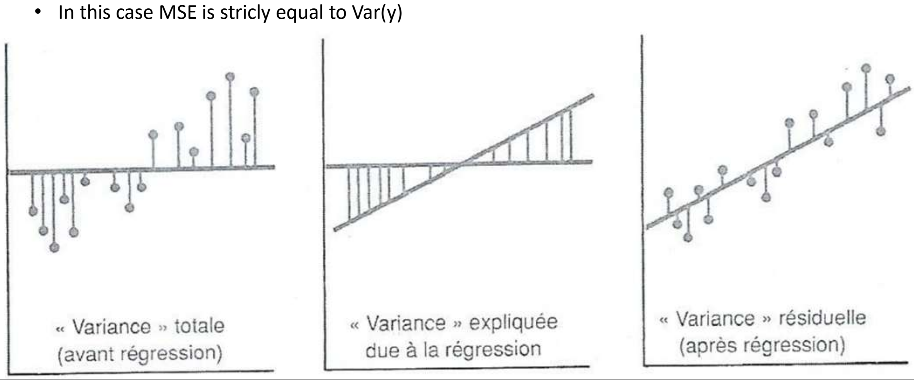
- MSE Comparison between
  - Regression model
  - Baseline inept model
    - Predict nothing but the mean : ŷ always = $\bar{y}$
    - In this case MSE is stricly equal to Var(y)

$$MSE = \frac{1}{n}\sum_{j=1}^{n}(y_j - \overset{\bar{y}}{\overset{\downarrow}{\hat{y}_j}})^2 \xleftrightarrow{\quad = \quad} Var(y) = \frac{1}{n}\sum_{j=1}^{n}(y_j - \bar{y})^2$$

Characteristics

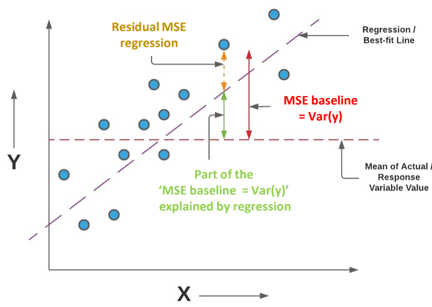


- MSE Comparison between
  - Regression model
  - Baseline inept model
    - Predict nothing but the mean : ŷ always = $\bar{y}$
    - In this case MSE is stricly equal to Var(y)



Characteristics

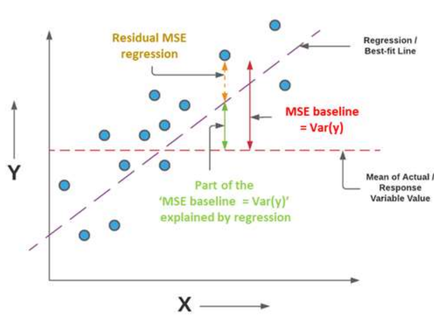
- A score (not an error)
  - The higher the best
  - Range between
    - 0 (0% of variance explained)
    - 1 (100% of variance explained)



R-square or determination coefficient

- Proportion or % of variance explained

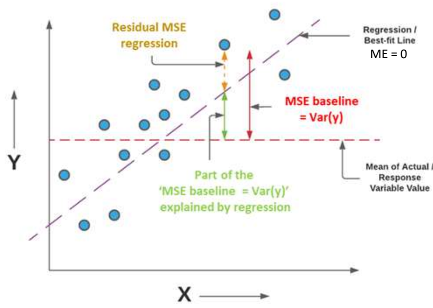
$$R^2 = 1 - \frac{\frac{\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2}{n}}{\frac{\sum_{j=1}^{n}\left(y_j - \bar{y}\right)^2}{n}} = \underbrace{1 - \frac{\overbrace{residual\ MSE}^{\text{Proportion of variance not explained by model}}}{baseline\ MSE\ or\ Var(y)}}_{\text{Complementary proportion (1-), i.e. Proportion of variance explained by model}}$$



Sklearn's 'explained_variance_score'

- Very similar to R^2
  - Same but Mean Error substract from residual MSE-

$$\textit{explained variance score} = 1 - \frac{\frac{\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2 - \textit{Mean Error}}{n}}{\frac{\sum_{j=1}^{n}\left(y_j - \bar{y}\right)^2}{n}}$$

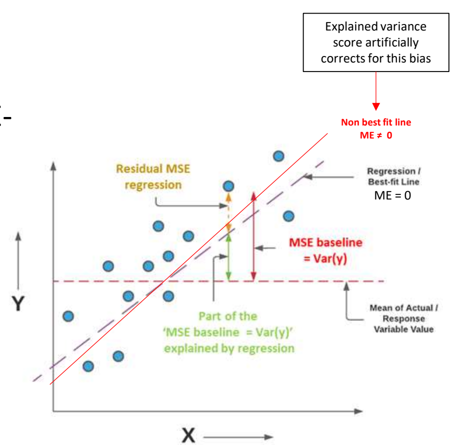


  - If best-fit line reached, ME always = 0
    - Thus, almost always R^2 = 'explained variance score'
  - Usefulness ?

Sklearn's 'explained_variance_score'

- Very similar to R^2
  - Same but Mean Error substract from residual MSE-

$$\textit{explained variance score} = 1 - \frac{\frac{\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2 - \textit{Mean Error}}{n}}{\frac{\sum_{j=1}^{n}\left(y_j - \bar{y}\right)^2}{n}}$$



  - If best-fit line reached, ME always = 0
    - Thus, almost always R^2 = 'explained variance score'
  - Usefulness ?
    - When estimator is biased (model not yet converged)
      - If R^2 = 'explained variance score' → convergence reached
      - If R^2 < 'explained variance score' → convergence not yet reached

Sklearn

```
Entrée [46]: from sklearn.metrics import explained_variance_score, r2_score

Entrée [47]: explained_variance_score(y_ess1, y_pred1)
    Out[47]: 0.9664748981730923

Entrée [48]: r2_score(y_ess1, y_pred1)
    Out[48]: 0.9664748981730923
```