机器学习基础 | 回归模型评估指标

原创 AhongPlus dataxon 2019-06-10

收录于话题
#机器学习 4个

回归模型中常用的评估指标可以分如下几类：

1. MAE系列，即由Mean Absolute Error衍生得到的指标；
2. MSE系列，即由Mean Squared Error衍生得到的指标；
3. R^2系列；

注：在英语中，error和deviation的含义是一样的，所以Mean Absolute Error也可以叫做Mean Absolute Deviation(MAD)，其他指标同理可得；

1 MAE系列

MAE全称Mean Absolute Error(平均绝对误差)。

更多参考：https://en.wikipedia.org/wiki/Mean_absolute_error

设N为样本数量，y_i为实际值，y'_i为预测值，那么MAE的定义如下

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - y'_i|$$

由MAE衍生可以得到：

Mean Absolute Pencentage Error(MAPE，平均绝对百分比误差)，相当于加权版的MAE.

$$MAPE=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i-y_i'}{y_i}\right|$$

MAPE 可 以 看 做 是 MAE 和 MPE(Mean Percentage Error) 综 合 而 成 的 指 标 $MPE=\frac{100\%}{N}\sum_{i=1}^{N}\frac{y_i-y_i'}{y_i}$.

从MAPE公式中可以看出有个明显的bug——当实际值y_i为0时就会得到无穷大值(实际值y_i的绝对值<1也会过度放大误差)。为了避免这个bug，MAPE一般用于实际值不会为0的情形。

Sungil Kima & Heeyoung Kim(2016)提出MAAPE(mean arctangent absolute percentage error)方法，在保持MAPE的算法思想下克服了上面那个bug(更多参考 A new metric of absolute percentage error for intermittent demand forecasts,Sungil Kima & Heeyoung Kim, 2016).

$$MAAPE=\frac{1}{N}\sum_{i=1}^{N}arctan\left(\left|\frac{y_i-y_i'}{y_i}\right|\right)$$

考虑Absolute Error $|y_i-y_i'|$可能存在Outlier的情况，此时Median Abosulte Error(MedAE, 中位数绝对误差)可能是更好的选择。

$$MedAE=\underset{i=1,\ldots,N}{median}|y_i-y_i'|$$

2 MSE系列

MSE全称Mean Squared Error(均方误差)，也可以称为Mean Squared Deviation (MSD).更多参考：https://en.wikipedia.org/wiki/Mean_squared_error

$$MSE=\frac{1}{N}\sum_{i=1}^{N}|y_i-y_i'|^2$$

由MSE可以衍生得到均方根误差(Root Mean Square Error, RMSE, 或者RMSD)更多参考：https://en.wikipedia.org/wiki/Root-mean-square_deviation

$$RMSE=\sqrt{MSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}|y_i-y_i'|^2}$$

RMSE可以进行归一化(除以全距或者均值)从而得到归一化的均方根误差(Normalized Root Mean Square Error, NRMSE).

$$NRMSE=\frac{RMSE}{y_{max}-y_{min}}$$

$$NRMSE=\frac{RMSE}{\bar{y}}$$

RMSE还有其他变式：

- RMSLE(Root Mean Square Logarithmic Error)

$$RMSLE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|log(y_i+1)-log(y_i'+1)|^2}$$

- RMSPE(Root Mean Square Percentage Error) $RMSPE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i-y_i'}{y_i}\right|^2}$

对于数值序列出现长尾分布的情况，可以选择MSLE(Mean squared logarithmic error，均方对数误差)，对原有数据取对数后再进行比较(公式中+1是为了避免数值为0时出现无穷值).

$$MSLE = \frac{1}{N}\sum_{i=1}^{N}|log(y_i+1)-log(y_i'+1)|^2$$

3 R²系列

R²(R squared, Coefficient of determination)，中文翻译为“决定系数”或者“拟合优度”，反映的是预测值对实际值的解释程度.

注意：R²和相关系数的平方不是一回事(只在简单线性回归条件下成立)，

$$\begin{aligned}R^2 &= 1-\frac{SS_{res}}{SS_{tot}}\\ &= 1-\frac{\sum_{i=1}^{N}(y_i-y_i')^2}{\sum_{i=1}^{N}(y_i-\overline{y})^2}\end{aligned}$$

其中$\overline{y}=\frac{1}{n}\sum_{i=1}^{N}y_i$，总平方和($SS_{tot}$) = 回归平方和($SS_{reg}$)+残差平方和(SS_{res}).

$$SS_{tot}=\sum_{i=1}^{N}(y_i-y_i')^2$$
$$SS_{res}=\sum_{i=1}^{N}(y_i-\overline{y})^2$$
$$SS_{reg}=\sum_{i=1}^{N}(y_i'-\overline{y})^2)$$
.

回归模型中，增加额外的变量会提升R²，但这种提升可能是虚假的，因此提出矫正的R²(Adjusted R²，符号表示为R^2_{adj}或$\overline{R}^2$)来对模型中的变量个数进行“惩罚”($R^2_{adj}\leq R^2$)。

$$R^2_{adj}=1-(1-R^2)\frac{N-1}{N-1-P}$$

公式中P表示回归模型中变量(特征)的个数。

和R²计算方式很相近的另一个指标是Explained Variance Score.

设$e_i=y_i-y_i',\overline{e}=\frac{1}{N}\sum_{i=1}^{N}e_i$，则有

$$
\begin{aligned}
explained\_variance &= 1 - \frac{var(y-y')}{var(y)} \\
&= 1 - \frac{\sum_{i=1}^{N}(e_i - \overline{e})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2}
\end{aligned}
$$

更多关于R^2参考：

- https://en.wikipedia.org/wiki/Coefficient_of_determination
- https://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit
- https://www.displayr.com/8-tips-for-interpreting-r-squared/

综上，在选用评价指标时，需要考虑

- 数据中是否有0，如果有0值就不能用MPE、MAPE之类的指标；
- 数据的分布如何，如果是长尾分布可以选择带对数变换的指标，中位数指标比平均数指标更好；
- 是否存在极端值，诸如MAE、MSE、RMSE之类容易受到极端值影响的指标就不要选用；
- 得到的指标是否依赖于量纲(即绝对度量，而不是相对度量)，如果指标依赖量纲那么不同模型之间可能因为量纲不同而无法比较；

更多关于指标选择可以参考*A Survey of Forecast Error Measures*(2013)这篇文章。

参考资料：

- *https://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/*
- A Survey of Forecast Error Measures, 2013
- *http://www.damienfrancois.be/blog/files/modelperfcheatsheet.pdf*
- *https://scikit-learn.org/stable/modules/model_evaluation.html*
- A new metric of absolute percentage error for intermittent demand forecasts,Sungil Kima & Heeyoung Kim, 2016
- Accuracy in forecasting: A survey, Essam Mahmoud, 1984