Review of Some Accuracy Metrics

1. Common Evaluation Metrics

1.1 Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of prediction accuracy of a forecasting method in statistics; for example in trend estimation, it's also used as a loss function for regression problems in machine learning [1]. The formula is shown below; where A_t is the actual value and F_t is the forecast value.

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|,$$

The absolute value in this calculation is summed for every forecasted point in time and divided by the number of fitted points n. Multiplying by 100% makes it a percentage error. Mean absolute percentage error is commonly used as a loss function for regression problems and in model evaluation, because of its very intuitive interpretation in terms of relative error. The mean absolute percentage error (MAPE) is the most common measure used to forecast error, and works best if there are no extremes to the data (and no zeros) [2].

1.2 Mean Absolute Error (MAE)

The mean absolute error is an average of the absolute errors. In statistics, mean absolute error (MAE) is a measure of difference between two continuous variables. Assume x and x_i are variables of paired observations that express the same phenomenon [3]. The formula for calculating MAE is shown below;

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - x|$$

The steps are relatively simple [4];

- Find all of your absolute errors in the dataset (x_i - x)
- Add them all up
- Divide by the number of errors. For example, if you had 10 measurements, divide by 10.

1.3 Mean Percent Error (MPE)

In statistics, the mean percentage error (MPE) is the computed average of percentage errors by which forecasts of a model differ from actual values of the quantity being forecast. The formula is given by;

$$\text{MPE} = \frac{100\%}{n}\sum_{t=1}^{n}\frac{a_t - f_t}{a_t}$$

where a_t is the actual value of the quantity being forecast, f_t is the forecast, and n is the number of different times for which the variable is forecast. As opposed to the MAPE; actual values rather than absolute values of the forecast errors are used in the formula, positive and negative forecast errors can offset each other; as a result the formula can be used as a measure of the bias in the forecasts [5].

1.4 Root Mean Squared Error (RMSE)

The root mean squared error (rmse) also known as the root mean square deviation (rmsd), is a frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed. RMSE is a measure of accuracy, to compare forecasting errors of different models for a particular dataset and not between datasets, as it is scale-dependent. RMSE is always non-negative, and a value of would indicate a perfect fit to the data. In general, a lower RMSE value is better than a higher one.

The RMSE value for a dataset is calculated by;

- Calculating the errors between the forecasts and the actuals

- Squaring the errors
- Finding the average of the residuals
- Taking the square root of the result

The formula for the root mean squared error can be seen below;

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

2. Probabilistic Forecasting Metrics

Probabilistic forecasts assign a probability to every possible future. A probability forecast refers to a specific event, such as there is a 25% probability of it raining in the next 24 hours. All probabilistic forecasts are not equally accurate, and metrics are needed to assess the respective accuracy of distinct probabilistic forecasts. Simple accuracy metrics such as MAE (Mean Absolute Error) or MAPE (Mean Absolute Percentage Error) are not directly applicable to probabilistic forecasts [6].

2.1 Brier Score

A brier score is a way to verify the accuracy of a probability forecast. The Brier score can only be used for binary outcomes, where there are only two possible events, like “it rained” or “it didn’t rain.” It could also be used for categorical outcomes as long as they can be structured as binary outcomes (i.e. “true” or “false”) [7]. The Brier score measures the mean squared difference between; the predicted probability assigned to the possible outcomes for an item, and the actual outcome. The best possible Brier score is 0 which means total accuracy; while the worst possible score is 1, which means the forecast was wholly inaccurate. The most common formula for Brier score is shown below (it’s similar to the MSE) [8];

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2$$

Where:

- N = the number of items you're calculating a Brier score for.
- f_t is the forecast probability,
- o_t is the outcome (1 if it happened, 0 if it didn't).

2.2 Continuous Ranked Probability Score (CRPS)

The CPRS is one of the most widely used accuracy metrics where probabilistic forecasts are involved. The Continuous Ranked Probability Score (CRPS) generalizes the MAE to the case of probabilistic forecasts. The CRPS is frequently used in order to assess the respective accuracy of two probabilistic forecasting models. The CRPS metric can be combined with a backtesting process in order to stabilize the accuracy assessment by leveraging multiple measurements over the same dataset.

The CRPS metric differs from simpler metrics (e.g., MAE) because of its asymmetric expression; the observations are deterministic, while the forecasts are probabilistic. The CPRS does not focus on any specific point of the probability distribution, but considers the distribution of the forecasts as a whole.

The CRPS metric can be defined with following steps;

- Let X be a random variable.
- Let F be the cumulative distribution function (CDF) of X, such as $F(y) = P[X \leq y]$.
- Let x be the observation, and F the CDF associated with an empirical probabilistic forecast.
- The CRPS between x and F is can be defined as:

$$CRPS(F, x) = \int_{-\infty}^{\infty} \Big(F(y) - \mathbb{1}(y - x)\Big)^2 dy$$

where $\mathbb{1}$ is the Heaviside step function and denotes a step function along the real line that attains:

- the value of 1 if the real argument is positive or zero,
- the value of 0 otherwise.

The CRPS is expressed in the same unit as the observed variable. The CRPS generalizes the mean absolute error; in fact, it reduces to the mean absolute error (MAE) if the forecast is deterministic. From a numerical perspective, a simple way of computing CPRS consists of breaking down the original integral into two integrals on well-chosen boundaries to simplify the Heaviside step function, which gives:

$$CRPS(F, x) = \int_{-\infty}^{x} F(y)^2 dy + \int_{x}^{\infty} \Big(F(y) - 1\Big)^2 dy$$

In practice, since F is an empirical distribution obtained through a forecasting model, the corresponding random variable X has a compact support, meaning that there is only a finite number of points where $P[X = x] > 0$. Thus, the integrals can be turned into discrete finite sums [6].

2.3 Cross – Entropy

In information theory, the cross entropy between two probability distributions p and q over the same underlying set of events measures the average number of bits needed to identify an event drawn from the set if a coding scheme used for the set is optimized for an estimated probability distribution q, rather than the true distribution p [9]. Cross-entropy is a measure of the difference between two probability distributions for a given random variable or set of events [10].

The cross-entropy has strong ties with the maximum likelihood estimation. Cross-entropy is of primary importance to modern forecasting systems, because if it is instrumental in making possible the delivery of superior forecasts, even for alternative metrics.

For two discrete random variables p and q, the cross-entropy is defined as:

$$H(p, q) = -\sum_x p(x)\ \log q(x).$$

P is intended as the “true” distribution, only partially observed, while Q is intended as the “unnatural” distribution obtained from a constructed statistical model. In information theory, cross-entropy can be interpreted as the expected length in bits for encoding messages, when Q is used instead of P.

In practice, as P isn’t known, the cross-entropy is empirically estimated from the observations, by simply assuming that all the collected observations are equally probable, that is, p(x) = 1/N where N is the number of observations [11].

$$H(q) = -\frac{1}{N} \sum_x \log q(x).$$

Cross-entropy builds upon the idea of entropy from information theory and calculates the number of bits required to represent or transmit an average event from one distribution compared to another distribution.

3 References

[1] "Mean absolute percentage error - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Mean_absolute_percentage_error. [Accessed: 16-Feb-2020].

[2] "Mean absolute percentage error (MAPE) - Statistics How To." [Online]. Available: https://www.statisticshowto.datasciencecentral.com/mean-absolute-percentage-error-mape/. [Accessed: 16-Feb-2020].

[3] "Mean absolute error - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Mean_absolute_error. [Accessed: 17-Feb-2020].

[4] "Absolute Error & Mean Absolute Error (MAE) - Statistics How To." [Online]. Available: https://www.statisticshowto.datasciencecentral.com/absolute-error/. [Accessed: 17-Feb-2020].

[5] "Mean percentage error - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Mean_percentage_error. [Accessed: 17-Feb-2020].

[6] "Continuous Ranked Probability Score (CRPS)." [Online]. Available: https://www.lokad.com/continuous-ranked-probability-score. [Accessed: 27-Feb-2020].

[7] "Brier Score: Definition, Examples - Statistics How To." [Online]. Available: https://www.statisticshowto.datasciencecentral.com/brier-score/. [Accessed: 27-Feb-2020].

[8] "Brier score - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Brier_score. [Accessed: 27-Feb-2020].

[9] "Cross entropy - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Cross_entropy. [Accessed: 27-Feb-2020].

[10] "A Gentle Introduction to Cross-Entropy for Machine Learning." [Online]. Available: https://machinelearningmastery.com/cross-entropy-for-machine-learning/. [Accessed: 27-Feb-2020].

[11] "Cross-Entropy Definition." [Online]. Available: https://www.lokad.com/cross-entropy-definition. [Accessed: 27-Feb-2020].