MSE

Среднеквадратичная ошибка (Mean Squared Error) применяется в случаях, когда требуется подчеркнуть большие ошибки и выбрать модель, которая дает меньше именно больших ошибок. Большие значения ошибок становятся заметнее за счет квадратичной зависимости.

Действительно, допустим модель допустила на двух примерах ошибки 5 и 10. В абсолютном выражении они отличаются в два раза, но если их возвести в квадрат, получив 25 и 100 соответственно, то отличие будет уже в четыре раза. Таким образом модель, которая обеспечивает меньшее значение MSE допускает меньше именно больших ошибок.

MSE рассчитывается по формуле:

$$\text{MSE} = \frac{1}{n}(y_i - \hat{y}_i)^2$$

где n – количество наблюдений по которым строится модель и количество прогнозов, y_i – фактические значение зависимой переменной для i-го наблюдения, $\hat{y}_i$ – значение зависимой переменной, предсказанное моделью.

Таким образом, можно сделать вывод, что MSE настроена на отражение влияния именно больших ошибок на качество модели.

Недостатком использования MSE является то, что если на одном или нескольких неудачных примерах, возможно, содержащих аномальные значения будет допущена значительная ошибка, то возведение в квадрат приведёт к ложному выводу, что вся модель работает плохо. С другой стороны, если модель даст небольшие ошибки на большом числе примеров, то может возникнуть обратный эффект — недооценка слабости модели.

RMSE

Корень из среднеквадратичной ошибки (Root Mean Squared Error) вычисляется просто как квадратный корень из MSE (среднеквадратичная ошибка):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

MSE и RMSE могут минимизироваться с помощью одного и того же функционала, поскольку квадратный корень является неубывающей функцией. Например, если у нас есть два набора результатов работы модели, A и B, и MSE для A больше, чем MSE для B, то мы можем быть уверены, что RMSE для A

больше RMSE для B. Справедливо и обратное: если $MSE(A) < MSE(B)$, то и $RMSE(A) < RMSE(B)$.

Следовательно, сравнение моделей с помощью RMSE даст такой же результат, что и для MSE. Однако с MSE работать несколько проще, поэтому она более популярна у аналитиков. Кроме этого, имеется небольшая разница между этими двумя ошибками при оптимизации с использованием градиента:

$$\frac{\partial RMSE}{\partial \hat{y}_i} = \frac{1}{2\sqrt{MSE}} \frac{\partial MSE}{\partial \hat{y}_i}$$

Это означает, что перемещение по градиенту MSE эквивалентно перемещению по градиенту RMSE, но с другой скоростью, и скорость зависит от самой оценки MSE. Таким образом, хотя RMSE и MSE близки с точки зрения оценки моделей, они не являются взаимозаменяемыми при использовании градиента для оптимизации.

Влияние каждой ошибки на RMSE пропорционально величине квадрата ошибки. Поэтому большие ошибки оказывают непропорционально большое влияние на RMSE. Следовательно, RMSE можно считать чувствительной к аномальным значениям.

MAE

Средняя абсолютная ошибка (Mean Absolute Error) вычисляется следующим образом:

$$\mathrm{M}A\mathrm{E} = \frac{1}{n}\sum_{i=1}^{n} |y_i \; - \; \hat{y}_i|$$

Т.е. MAE рассчитывается как среднее абсолютных разностей между наблюдаемым и предсказанным значениями. В отличие от MSE и RMSE она является линейной оценкой, а это значит, что все ошибки в среднем взвешены одинаково. Например, разница между 0 и 10 будет вдвое больше разницы между 0 и 5. Для MSE (среднеквадратичная ошибка) и RMSE (корень из среднеквадратичной ошибки), как отмечено выше, это не так.

Поэтому MAE широко используется, например, в финансовой сфере, где ошибка в 10 долларов должна интерпретироваться как в два раза худшая, чем ошибка в 5 долларов.