

Анализ с помощью характеристической кривой (ROC- analysis)

Синонимы: ROC-анализ

Разделы: [Визуализация](#), [Алгоритмы](#)

ROC-анализ представляет собой графический метод оценки качества работы бинарного классификатора и выбора дискриминационного порога для разделения классов. В основе метода лежит построение ROC-кривой (ROC — receiver operating characteristic — рабочая характеристика приемника) — графического представления зависимости двух величин: чувствительности и специфичности.

Под чувствительностью (Se — sensitivity) бинарной модели понимается доля истинно-положительных классификаций в общем числе положительных наблюдений (TPR — true-positive rate):

$$Se = TPR = \frac{TP}{P} = \frac{TP}{TP+FN}$$

Это есть доля правильно классифицированных положительных наблюдений. Следовательно, чем выше чувствительность, тем более надежно классификатор распознает положительные примеры.

Под специфичностью (Sp — specify) модели понимается доля истинно-отрицательных классификаций в общем числе отрицательных наблюдений (TNR — true-negative rate):

$$Sp = TNR = \frac{TN}{N} = \frac{TN}{TN+FP}$$

Таким образом, чем выше специфичность, тем более надежно классификатор распознает отрицательные наблюдения.

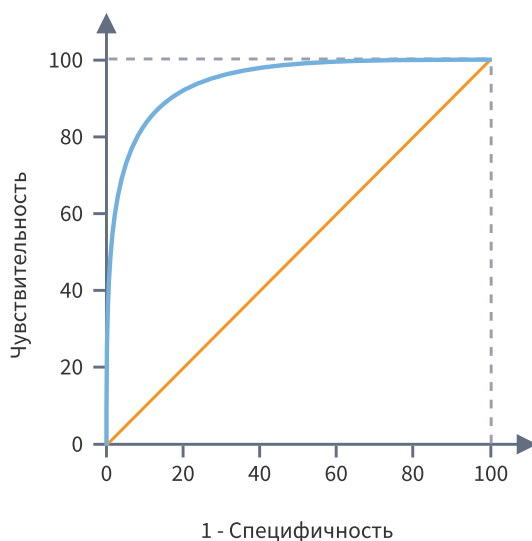
По вертикальной оси графика ROC-кривой представлена чувствительность, а по горизонтальной — величина, равная $1 - Sp$. Можно показать, что

$$1 - Sp = FPR = \frac{FP}{N} = \frac{FP}{TN+FP}$$

Это доля ложно-положительных классификаций в общем числе отрицательных классификаций, которую можно интерпретировать как вероятность «ложной тревоги».

Таким образом, ROC-кривая отражает связь между вероятностью ложной тревоги (доли ложно-положительных классификаций) и вероятностью «правильного обнаружения» (доли истинно-положительных классификаций). С ростом чувствительности растёт

надежность распознавания положительных наблюдений (снижается вероятность «пропуска цели»), но при этом растет вероятность ложной тревоги.



На рисунке линии $[(0,0); (0, 100)]$ и $[(0,100); (100, 100)]$ образуют ROC-кривую идеального классификатора, когда рост чувствительности (надежности классификации) вообще не сопровождается ростом вероятности ложной тревоги. Это мало реальный случай, когда точность классификации положительных наблюдений вообще не зависит от уровня дискриминационного порога.

Обычно ROC-кривая реальной модели имеет параболическую форму с различной кривизной. При этом, чем выше кривизна и кривая ближе к идеальной, тем лучше работает модель. При ухудшении качества модели ROC-кривая вырождается в диагональную прямую линию, которая соответствует «бесполезному» классификатору, который предсказывает классы случайным образом.

ROC-анализ применяется для оценки качества моделей: позволяет аналитику выбрать модель с наилучшей прогностической силой, проанализировать чувствительность и специфичность моделей, подобрать порог отсечения. Подробнее в статье [«Логистическая регрессия и ROC-анализ — математический аппарат»](#).

В Loginom существует специализированный визуализатор [качество бинарной классификации](#), в котором строятся диаграммы (в том числе ROC-кривая) и таблицы с результатами проведенной классификации на основе [логистической регрессии](#). А решение [Loginom Scorecard Modeler](#), автоматизирующее процесс построения скоринговых карт, включает оценку качества модели при помощи ROC-кривой.