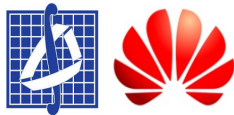


Введение в искусственный интеллект. Машинное обучение

Тема семинара: метрики качества классификаторов

Бабин Д.Н., Иванов И.Е.

кафедра Математической Теории Интеллектуальных Систем



① Матрица ошибок

- 1 Матрица ошибок
- 2 Ошибки на основе положительного отклика (TPR, FPR) и площадь под кривой ROC

- 1 Матрица ошибок
- 2 Ошибки на основе положительного отклика (TPR, FPR) и площадь под кривой ROC
- 3 Точность, полнота (Precision, Recall) и площадь под кривой PR

- 1 Матрица ошибок
- 2 Ошибки на основе положительного отклика (TPR, FPR) и площадь под кривой ROC
- 3 Точность, полнота (Precision, Recall) и площадь под кривой PR
- 4 Микро- и макроусреднение

Классификация ответов бинарного классификатора

- Обучающая выборка $X^m = \{(x_1, y_1), \dots, (x_m, y_m)\}$
- Задача классификации на 2 класса: $X \rightarrow Y, Y = \{+1, -1\}$
- Алгоритм классификации $a(x) : X \rightarrow Y$
- Класс с меткой “+1” называется “**positive**”
- Класс с меткой “-1” называется “**negative**”



Классификация ответов бинарного классификатора

- Обучающая выборка $X^m = \{(x_1, y_1), \dots, (x_m, y_m)\}$
- Задача классификации на 2 класса: $X \rightarrow Y, Y = \{+1, -1\}$
- Алгоритм классификации $a(x) : X \rightarrow Y$
- Класс с меткой “+1” называется “**positive**”
- Класс с меткой “-1” называется “**negative**”

Таблица: Классификация ответов

	Выход алгоритма	Правильный ответ
TP (True Positive)	$a(x_i) = +1$	$y_i = +1$
TN (True Negative)	$a(x_i) = -1$	$y_i = -1$
FP (False Positive)	$a(x_i) = +1$	$y_i = -1$
FN (False Negative)	$a(x_i) = -1$	$y_i = +1$


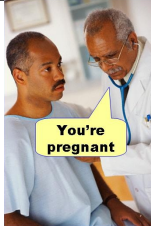




Матрица ошибок

Более наглядно эти соотношения можно изобразить с помощью **матрицы ошибок (confusion matrix)**

		Правильный ответ	
		$y = +1$	$y = -1$
Выход алгоритма	$a(x) = +1$	True Positive	False Positive (Ошибка 1 рода)
	$a(x) = -1$	False Negative (Ошибка 2 рода)	True Negative

Матрица ошибок

	$y = +1$	$y = -1$
$a(x) = +1$		
$a(x) = -1$		

Простейшая метрика качества

- Простейшая метрика качества - это доля правильных ответов на тесте (контрольной выборке)
- По-английски - **Accuracy**

Формула Accuracy

$$Accuracy = \frac{1}{m} \sum_{i=1}^m [a(x_i) = y_i] = \frac{TP+TN}{TP+FP+TN+FN}$$



Простейшая метрика качества

- Простейшая метрика качества - это доля правильных ответов на тесте (контрольной выборке)
- По-английски - **Accuracy**

Формула Accuracy

$$Accuracy = \frac{1}{m} \sum_{i=1}^m [a(x_i) = y_i] = \frac{TP+TN}{TP+FP+TN+FN}$$

Недостаток

- Не учитывается дисбаланс классов
- Не учитывается цена ошибки на объектах разных классов



Метрики по положительному отклику алгоритма

Рассмотрим метрики, которые основаны на подсчёте доли положительных ответов алгоритма.

Доля ложных положительных классификаций

Также известно как False Positive Rate, или **FPR**.

$$FPR(a, X^m) = \frac{\sum_{i=1}^m [y_i = -1][a(x_i) = +1]}{\sum_{i=1}^m [y_i = -1]}$$

Метрики по положительному отклику алгоритма

Рассмотрим метрики, которые основаны на подсчёте доли положительных ответов алгоритма.

Доля ложных положительных классификаций

Также известно как False Positive Rate, или **FPR**.

$$FPR(a, X^m) = \frac{\sum_{i=1}^m [y_i = -1][a(x_i) = +1]}{\sum_{i=1}^m [y_i = -1]}$$

Доля верных положительных классификаций

Также известно как True Positive Rate, или **TPR**.

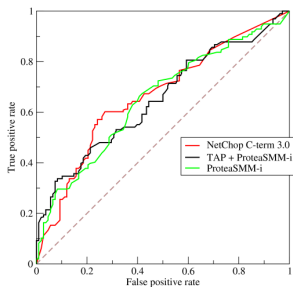
$$TPR(a, X^m) = \frac{\sum_{i=1}^m [y_i = +1][a(x_i) = +1]}{\sum_{i=1}^m [y_i = +1]}$$

Замечание. Обратите внимание на разные знаменатели!



Кривая ошибок

Наиболее известна как рабочая характеристика приёмника, или Receiver Operating Characteristic (**ROC-кривая**), в который мы смотрим на компромисс между уровнем ложной тревоги и долей верного отклика.



По оси X откладывается FPR, по оси Y - TPR¹.

Замечание. На данной кривой никак не учитываются пропуски.

¹<https://wikipedia.org>

AUROC

Чем больше для каждого значения ошибки FPR значение правильного предсказания TPR, тем лучше работает классификатор.

Т.о., площадь под кривой (Area Under Curve, AUC / AUROC) необходимо максимизировать.

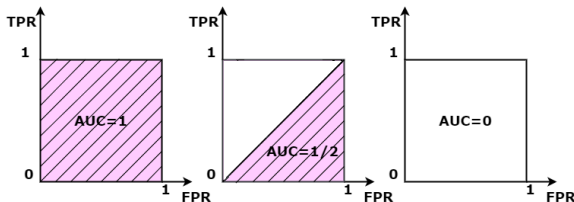
Площадь под ROC-кривой и виды ROC-кривых

AUROC

Чем больше для каждого значения ошибки FPR значение правильного предсказания TPR, тем лучше работает классификатор.

Т.о., площадь под кривой (Area Under Curve, AUC / AUROC) необходимо максимизировать.

Наглядны ROC-кривые для наилучшего ($AUC=1$), случайного ($AUC=0.5$) и наихудшего ($AUC=0$) алгоритма.



Задача

Предположим, что алгоритм бинарной классификации $a(x_i)$ на выборке X^m принимает решение о присвоении класса на основе некоторого скалярного значения $g_\theta(x_i) \in \mathbb{R}$, где θ - набор параметров модели, а $g_\theta(x_i)$ - дискриминантная функция:

- Положительный отклик по порогу t : $g_\theta(x_i) \geq t$

Задача

- Хотим построить ROC-кривую, т.е. найти точки $\{(FPR_i, TPR_i)\}_{i=1}^m$
- Подсчитать площадь под кривой - AUROC



Задача

Предположим, что алгоритм бинарной классификации $a(x_i)$ на выборке X^m принимает решение о присвоении класса на основе некоторого скалярного значения $g_\theta(x_i) \in \mathbb{R}$, где θ - набор параметров модели, а $g_\theta(x_i)$ - дискриминантная функция:

- Положительный отклик по порогу t : $g_\theta(x_i) \geq t$

Задача

- Хотим построить ROC-кривую, т.е. найти точки $\{(FPR_i, TPR_i)\}_{i=1}^m$
- Подсчитать площадь под кривой - AUROC

Подсчитаем количество правильных ответов разного типа:

- $m_+ = \sum_{i=1}^m [y(x_i) = +1]$ (знаменатель в TPR)
- $m_- = \sum_{i=1}^m [y(x_i) = -1]$ (знаменатель в FPR: $m = m_+ + m_-$)

Упорядочим обучающую выборку X^m по убыванию значений $g_\theta(x_i)$.

Тогда формула для $AUROC = \frac{1}{m_-} \sum_{i=1}^m [y_i = -1] TPR_i$ (см. далее).



Алгоритм

Первую точку ставим в начало координат: $(FPR_0, TPR_0) = (0, 0)$, $AUROC = 0$.

Алгоритм

Первую точку ставим в начало координат: $(FPR_0, TPR_0) = (0, 0)$, $AUROC = 0$.

Цикл по упорядоченной выборке $i = 1 \dots m$

Порог — очередное значение дискриминантной функции $t = g_\theta(x_i)$

Если $y_i = -1$:

- $(FPR_i, TPR_i) = (FPR_{i-1} + \frac{1}{m_-}, TPR_{i-1})$ (двигаемся по оси X)
- $AUROC = AUROC + \frac{1}{m_-} TPR_i$

Решение задачи

Алгоритм

Первую точку ставим в начало координат: $(FPR_0, TPR_0) = (0, 0)$, $AUROC = 0$.

Цикл по упорядоченной выборке $i = 1 \dots m$

Порог — очередное значение дискриминантной функции $t = g_\theta(x_i)$

Если $y_i = -1$:

- $(FPR_i, TPR_i) = (FPR_{i-1} + \frac{1}{m_-}, TPR_{i-1})$ (двигаемся по оси X)
- $AUROC = AUROC + \frac{1}{m_-} TPR_i$

Если $y_i = +1$:

- $(FPR_i, TPR_i) = (FPR_{i-1}, TPR_{i-1} + \frac{1}{m_+})$ (двигаемся по оси Y)

В задачах информационного поиска

- Точность, или $Precision = \frac{TP}{TP+FP}$ (доля релевантных объектов среди найденных)
- Полнота, или $Recall = \frac{TP}{TP+FN}$ (доля найденных объектов среди релевантных)

В задачах информационного поиска

- Точность, или $Precision = \frac{TP}{TP+FP}$ (доля релевантных объектов среди найденных)
- Полнота, или $Recall = \frac{TP}{TP+FN}$ (доля найденных объектов среди релевантных)

Как применяются

- **Точность:** позволяет следить, чтобы было мало ложных тревог; но при этом ничего не говорит о пропусках (высока цена ложной тревоги, а цена пропуска - низкая).
- **Полнота:** позволяет следить, чтобы было мало пропусков; но при этом ничего не говорит о ложных тревогах (высока цена пропуска, а цена ложной тревоги - низкая).

Замечание. Зачастую задача состоит в оптимизации одной метрики при фиксации другой.



В задачах медицинской диагностики

- Чувствительность, или $Sensitivity = \frac{TP}{TP+FN}$ (доля верных положительных диагнозов)
- Специфичность, или $Specificity = \frac{TN}{TN+FP}$ (доля верных отрицательных диагнозов)

В задачах медицинской диагностики

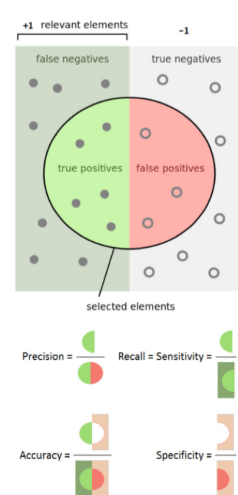
- Чувствительность, или $Sensitivity = \frac{TP}{TP+FN}$ (доля верных положительных диагнозов)
- Специфичность, или $Specificity = \frac{TN}{TN+FP}$ (доля верных отрицательных диагнозов)

Как применяются

- **Чувствительность:** максимизируем количество верных положительных диагнозов, но не учитываем ложные диагнозы (стоимость лечения низкая, а цена пропуска - высокая).
- **Специфичность:** максимизируем количество верных отрицательных диагнозов, но не учитываем пропуски диагноза (стоимость лечения высокая, а цена пропуска - низкая).

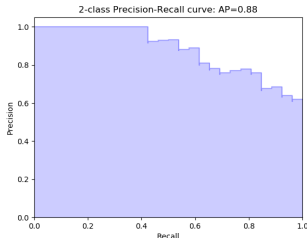


Иллюстрация метрик



Агрегированные метрики над Precision-Recall

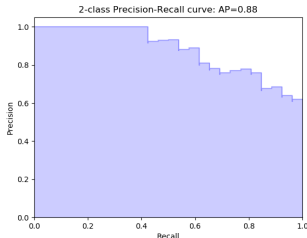
Можно построить кривую Точность-Полнота (PR-кривая) по аналогии с ROC-кривой:



Замечание. Обратите внимание, что в данном случае кривая не обязательно монотонна!

Агрегированные метрики над Precision-Recall

Можно построить кривую Точность-Полнота (PR-кривая) по аналогии с ROC-кривой:



Замечание. Обратите внимание, что в данном случае кривая не обязательно монотонна!

AUPRC

- Аналогично AUROC, можно вычислить площадь под PR-кривой - AUPRC
- Другое название - Average Precision (с некоторым допущениями на способ интегрирования): чем больше, тем лучше

Для каждого класса $c \in Y$ обозначим через TP_c , FP_c и FN_c верные положительные, ложные положительные и ложные отрицательные ответы. Тогда:

Точность и полнота с макроусреднением

- $Precision = \frac{\sum_c TP_c}{\sum_c (TP_c + FP_c)}$
- $Recall = \frac{\sum_c TP_c}{\sum_c (TP_c + FN_c)}$
- Не чувствительно к ошибкам на маленьких классах

Многоклассовая классификация

Для каждого класса $c \in Y$ обозначим через TP_c , FP_c и FN_c верные положительные, ложные положительные и ложные отрицательные ответы. Тогда:

Точность и полнота с макроусреднением

- $Precision = \frac{\sum_c TP_c}{\sum_c (TP_c + FP_c)}$
- $Recall = \frac{\sum_c TP_c}{\sum_c (TP_c + FN_c)}$
- Не чувствительно к ошибкам на маленьких классах

Точность и полнота с микроусреднением

- $Precision = \frac{1}{|Y|} \sum_c \frac{TP_c}{TP_c + FP_c}$
- $Recall = \frac{1}{|Y|} \sum_c \frac{TP_c}{TP_c + FN_c}$
- Чувствительно к ошибкам на маленьких классах



- Точность и полнота подходят для задач информационного поиска, когда доля объектов релевантного класса мала

Резюме по оценкам качества классификации

- Точность и полнота подходят для задач информационного поиска, когда доля объектов релевантного класса мала
- Чувствительность и специфичность подходят для задач с несбалансированными классами (как, например, в медицине)

Резюме по оценкам качества классификации

- Точность и полнота подходят для задач информационного поиска, когда доля объектов релевантного класса мала
- Чувствительность и специфичность подходят для задач с несбалансированными классами (как, например, в медицине)
- AUROC подходит для оценки качества при нефиксированном соотношении цены ошибок



Резюме по оценкам качества классификации

- Точность и полнота подходят для задач информационного поиска, когда доля объектов релевантного класса мала
- Чувствительность и специфичность подходят для задач с несбалансированными классами (как, например, в медицине)
- AUROC подходит для оценки качества при нефиксированном соотношении цены ошибок
- Ещё одна агрегированная оценка качества - F-мера:
$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$
 - Это *гармоническое среднее*, которое стремится к нулю когда хотя бы одно из значений стремится к нулю





На основе материалов сайта <http://www.machinelearning.ru>.