

# Оценка моделей

Виктор Владимирович Китов

МГУ им.Ломоносова, ф-т ВМиК, кафедра ММП.

I семестр 2015 г.

## Матрица ошибок (confusion matrix)

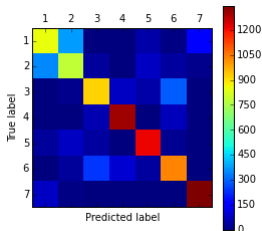
Матрица ошибок  $M = \{m_{ij}\}_{i,j=1}^C$  показывает количества объектов, реально принадлежащих классу  $\omega_i$ , но отнесенных к  $\omega_j$ .

		Прогнозный класс				
		1	2	...	C	
Истинный класс	1	[	$n_{11}$	$n_{12}$		
	2		$n_{21}$	$n_{22}$		
	⋮				⋮	
	C					$n_{CC}$
		]				

Диагональные элементы соответствуют правильным классификациям, а внедиагональные - неправильным.

# Пример визуализации матрицы ошибок

## Пример визуализации матрицы ошибок



- Видно, что ошибки сконцентрированы на разделении классов 1 и 2.
- Классы 1, 2 можно объединить в один класс «1+2», решить 6-то классовую задачу, а потом для всех объектов, отнесенных к «1+2», разделять их на классы 1 и 2 отдельным классификатором.

## Случай 2х классов

### Матрица ошибок:

		Прогноз	
		+	-
Правильный класс	+	TP (true positives)	FN (false negatives)
	-	FP (false positives)	TN (true negatives)

$P$  и  $N$  - число наблюдений положительного и отрицательного класса.

$$P = TP + FN, \quad N = TN + FP$$

## Меры качества

Точность (accuracy):	$\frac{TP+TN}{P+N}$
Доля ошибок (error rate):	$1-\text{accuracy}=\frac{FP+FN}{P+N}$
FPR (ложная тревога):	$\frac{FP}{N}$
TPR (вероятность обнаружения):	$\frac{TP}{P}$
Точность (precision):	$\frac{TP}{TP+FP}$
Полнота (recall):	$\frac{TP}{P}$
F-мера:	$\frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$
взвешенная F-мера:	$\frac{1}{\frac{\beta^2}{1+\beta^2} \frac{1}{Precision} + \frac{1}{1+\beta^2} \frac{1}{Recall}}$

## Разделимость и надежность

- **Разделимость (discriminability)** измеряет правильность соотнесения классов
  - например, все ранее перечисленные меры: доля ошибок, точность, полнота, и т.д.
- **Надежность (reliability)** измеряет правильность соотнесения вероятностей классов
  - Правдоподобие (что объект  $x_i$  принадлежит классу  $y_i$ ,  $i = 1, 2, \dots, N$ ):

$$\prod_{i=1}^N \hat{p}(y_i|x_i)$$

- Brier score:

$$\frac{1}{n} \sum_{i=1}^n \sum_{c=1}^C (\mathbb{I}[x_i \in \omega_c] - \hat{p}(\omega_c|x_i))^2$$

- Пример хорошей разделимости, но плохой надежности.

# Table of Contents

## 1 ROC кривые

## Байесовское решающее правило

- Обозначение:  $\hat{\omega}_i$  означает, что «прогноз равен классу  $\omega_i$ »
- Матрица цены:

	Прогноз	
	$\hat{\omega}_1$	$\hat{\omega}_2$
Правильный класс	$\omega_1$	0
	$\omega_2$	$\lambda_2$
		0

- $\lambda_1, \lambda_2$ -цена неправильной классификации класса  $\omega_1$  и  $\omega_2$  соответственно.



## Байесовское решающее правило

- Цена прогноза  $\hat{\omega}_1$ :  
 $L(\hat{\omega}_1) = \lambda_2 p(\omega_2|x) = \lambda_2 p(\omega_2)p(x|\omega_2)/p(x)$
- Цена прогноза  $\hat{\omega}_2$ :  
 $L(\hat{\omega}_2) = \lambda_1 p(\omega_1|x) = \lambda_1 p(\omega_1)p(x|\omega_1)/p(x)$
- *Байесовское правило* минимизирует ожидаемую цену:

$$\hat{\omega}^* = \arg \min_{\hat{\omega}} L(\hat{\omega})$$

- Оно эквивалентно:  
 $\hat{\omega}^* = \hat{\omega}_1 \Leftrightarrow \lambda_2 p(\omega_2)p(x|\omega_2) < \lambda_1 p(\omega_1)p(x|\omega_1) \Leftrightarrow$

$$\frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{\lambda_2 p(\omega_2)}{\lambda_1 p(\omega_1)} = \mu$$

## Дискриминативные решающие правила

- Правило, основанное на дискриминантных ф-циях:
  - соотнести  $x$  классу  $\omega_1 \iff g_1(x) - g_2(x) > \mu$
  - соотнести  $x$  классу  $\omega_1 \iff g_1(x)/g_2(x) > \mu$  (для  $g_1(x) > 0, g_2(x) > 0$ )
- Правило, основанное на вероятностях:
  - соотнести  $x$  классу  $\omega_1 \iff P(\omega_1|x) > \mu$

## ROC кривая

- ROC кривая - зависимость вероятности обнаружения положительного класса (TPR) от вероятности ложной тревоги (FPR) для различных  $\mu$ .
- Если  $\mu$  уменьшается, алгоритм чаще выбирает  $\omega_1$  и
  - TPR= $1 - \varepsilon_1$  возрастает
  - FPR= $\varepsilon_2$  также возрастает
- Диагональ соответствует случайной классификации  $\omega_1$  и  $\omega_2$  с вероятностями  $\mu$  и  $1 - \mu$ .
- Характеризует качество классификации для различных значений  $\mu$ .
  - более выпуклые вверх кривые лучше

## Изо-линии цены

- Обозначим  $\varepsilon_1, \varepsilon_2$  - вероятности ошибиться на классе  $\omega_1$  и  $\omega_2$  соответственно.
- $1 - \varepsilon_1 = TPR$ ,  $\varepsilon_2 = FPR$
- Ожидаемые потери

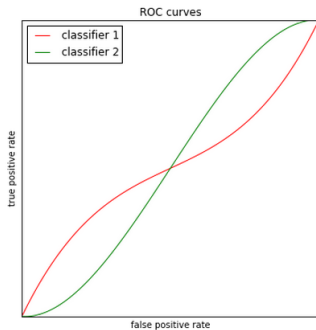
$$L = \lambda_2 p(\omega_2) \varepsilon_2 + \lambda_1 p(\omega_1) \varepsilon_1 = \lambda_2 p(\omega_2) \varepsilon_2 - \lambda_1 p(\omega_1) (1 - \varepsilon_1) + \lambda_1 p(\omega_1)$$

- Изо-линия потерь:

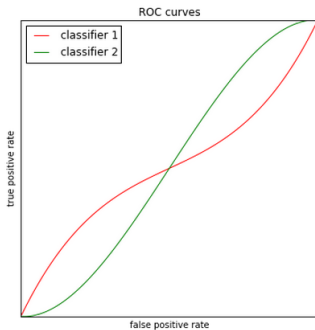
$$(1 - \varepsilon_1) = \frac{\lambda_2 p(\omega_2)}{\lambda_1 p(\omega_1)} \varepsilon_2 + \frac{\lambda_1 p(\omega_1) - L}{\lambda_1 p(\omega_1)}$$

- В оптимальной точке изо-линия касается ROC-кривой с тангенсом угла наклона  $\frac{\lambda_2 p(\omega_2)}{\lambda_1 p(\omega_1)}$

# Сравнение классификаторов по ROC кривой



# Сравнение классификаторов по ROC кривой



Как сравнивать различные классификаторы?

# Критерий AUC

- AUC - площадь под ROC-кривой:
  - глобальная характеристика качества
  - $AUC \in [0, 1]$ 
    - $AUC=0.5$  - эквивалент случайного угадывания
    - $AUC=1$  - безошибочное распознавание.
  - равна вероятности того, что для случайных  $x_1 \in \omega_1$  и  $x_2 \in \omega_2$  будет выполнено:  $\hat{p}(\omega_1|x_1) > \hat{p}(\omega_2|x)$

## LC index

- LC index - применение методики к байесовскому решающему правилу:
  - Отмасштабируем  $\lambda_1$  и  $\lambda_2$  так, что  $\lambda_1 + \lambda_2 = 1$
  - определим  $\lambda_1 = \lambda$ ,  $\lambda_2 = 1 - \lambda$
  - для каждого  $\lambda \in [0, 1]$  рассчитаем
$$L(\lambda) = \begin{cases} +1 & \text{если 1й классификатор лучше} \\ -1 & \text{если 2й классификатор лучше} \end{cases}$$
  - определим плотность вероятности  $\lambda$ :  $p(\lambda)$  (например, треугольник)
  - выбираем 1-й классификатор  $\iff \int_0^1 L(\lambda)p(\lambda)d\lambda > 0$ .



## Распределение вероятности ошибки

- Пусть  $e$  - вероятность ошибки на новом объекте.
- Цель - найти распределение вероятности  $e$ 
  - знаем, что на отложенной выборке объема  $n$  было  $k$  ошибок.

Вероятность сделать  $k$  ошибок на выборке объема  $n$ :

$$p(k|e, n) = \binom{n}{k} e^k (1 - e)^{n-k}$$

Тогда

$$p(e|k, n) = \frac{p(e, k|n)}{p(k|n)} = \frac{p(k|e, n)p(e|n)}{\int p(k|n)p(e|n)de}$$

Полагая априорное распределение  $p(e|n) \equiv \text{const}$ , получим

$$p(e|k, n) = \frac{p(k|e, n)}{\int p(k|n)de} \propto e^k (1 - e)^{n-k}$$

## Распределение вероятности ошибки

Поскольку бета-распределение имеет вид

$$Be(x|\alpha, \beta) = [\Gamma(\alpha + \beta)/(\Gamma(\alpha)\Gamma(\beta))]x^{\alpha-1}(1-x)^{\beta-1}, \text{ то}$$

$$p(e|k, n) \sim Be(k + 1, n - k + 1)$$

Бета-распределение:

$$\xi \sim Be(\alpha, \beta) \Rightarrow \mathbb{E}[\xi] = \frac{\alpha}{\alpha + \beta}, \text{ Var}[\xi] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

