

Оценка результатов анализа

Лекция 3

Метрики классификации

ID транзакции	Покупатель	Даты	Куплено фруктов	Куплена рыба	Потрачено
1	Пингвин	1 янв.	1	да	5,30 \$
2	Медведь	1 янв.	4	да	9,70 \$
3	Кролик	1 янв.	6	нет	6,50 \$
4	Лошадь	2 янв.	6	нет	5,50 \$
5	Пингвин	2 янв.	2	да	6,00 \$
6	Жираф	3 янв.	5	нет	4,80 \$
7	Кролик	3 янв.	8	нет	7,60 \$
8	Кот	3 янв.	?	да	7,40 \$

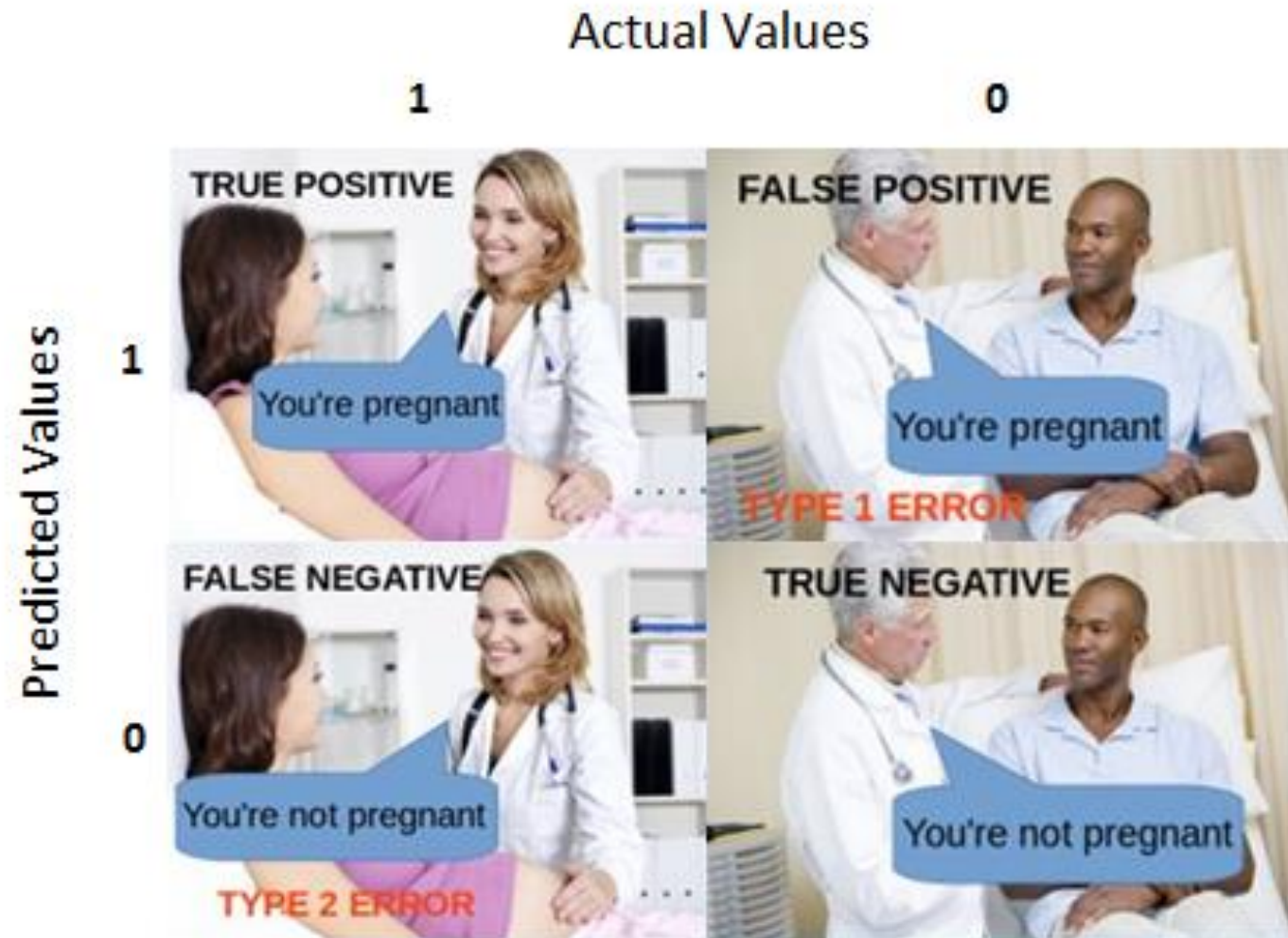
Метрики классификации

		Прогноз	
		Купят	Не купят
Факт	Купили	1 (TP)	5 (FN)
	Не купили	5 (FP)	89 (TN)

Матрица неточностей

		Прогноз	
		Купят	Не купят
Факт	Купили	1 (TP)	5 (FN)
	Не купили	5 (FP)	89 (TN)

Матрица неточностей

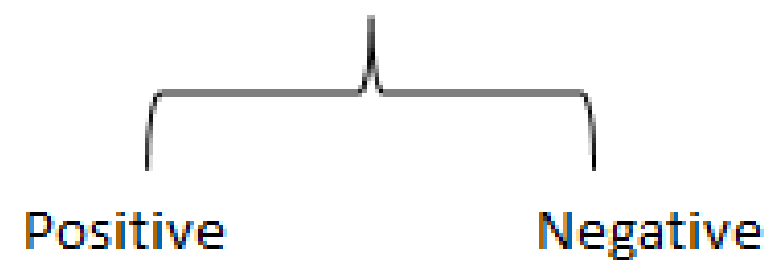


Матрица неточностей

Actual Values



Predicted Values



Матрица неточностей: как считать?

1. Каждая строка матрицы связана с прогнозируемым классом.
2. Каждый столбец матрицы соотносится с фактическим классом.
3. Общее количество правильных и неправильных классификаций заносится в таблицу.
4. Сумма правильных предсказаний для класса попадает в столбец предсказаний и строку ожидаемых значений для этого класса.
5. Сумма неправильных прогнозов для класса попадает в ожидаемую строку для данного значения класса и в прогнозируемый столбец для данного конкретного значения класса.

Матрица неточностей: пример

ID транзакции	Покупатель	Даты	Куплено фруктов	Куплена рыба	Потрачено
1	Пингвин	1 янв.	1	да	5,30 \$
2	Медведь	1 янв.	4	да	9,70 \$
3	Кролик	1 янв.	6	нет	6,50 \$
4	Лошадь	2 янв.	6	нет	5,50 \$
5	Пингвин	2 янв.	2	да	6,00 \$
6	Жираф	3 янв.	5	нет	4,80 \$
7	Кролик	3 янв.	8	нет	7,60 \$
8	Кот	3 янв.	?	да	7,40 \$

Матрица неточностей: пример

id транзакции	Покупатель	Даты	Куплено фруктов	Куплена рыба	Потрачено	Vegetarian	
0	1	Пингвин	01.янв	1.0	да	5,3	False
1	2	Медведь	01.янв	4.0	да	9,7	False
2	3	Кролик	01.янв	6.0	нет	6,5	True
3	4	Лошадь	02.янв	6.0	нет	5,5	True
4	5	Пингвин	02.янв	2.0	да	6	False
5	6	Жираф	03.янв	5.0	нет	4,8	True
6	7	Кролик	03.янв	8.0	нет	7,6	True

1. Не вегетарианцы покупают рыбу

2. Те, кто покупает мало фруктов, покупают рыбу

id транзакции	Покупатель	Даты	Куплено фруктов	Куплена рыба	Потрачено	Vegetarian	LessFruit	
0	1	Пингвин	01.янв	1.0	да	5,3	False	True
1	2	Медведь	01.янв	4.0	да	9,7	False	False
2	3	Кролик	01.янв	6.0	нет	6,5	True	False
3	4	Лошадь	02.янв	6.0	нет	5,5	True	False
4	5	Пингвин	02.янв	2.0	да	6	False	True
5	6	Жираф	03.янв	5.0	нет	4,8	True	False
6	7	Кролик	03.янв	8.0	нет	7,6	True	False

Матрица неточностей: пример

1. Для случая с не-вегетарианцами получаем такую матрицу:

```
confusion_matrix(y_true, y_pred_veg)  
array([[4, 0],  
       [0, 3]], dtype=int64)
```

2. Для случая с покупкой фруктов - такую:

```
confusion_matrix(y_true, y_pred_fru)  
array([[4, 0],  
       [1, 2]], dtype=int64)
```

Матрица неточностей: пример

1. Для случая с не-вегетарианцами получаем такую матрицу:

```
confusion_matrix(y_true, y_pred_veg)  
array([[4, 0],  
       [0, 3]], dtype=int64)
```

2. Для случая с покупкой фруктов - такую:

```
confusion_matrix(y_true, y_pred_fru)  
array([[4, 0],  
       [1, 2]], dtype=int64)
```

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Точность и полнота

Точность показывает, какая доля объектов, выделенных классификатором как положительные, действительно является положительными:

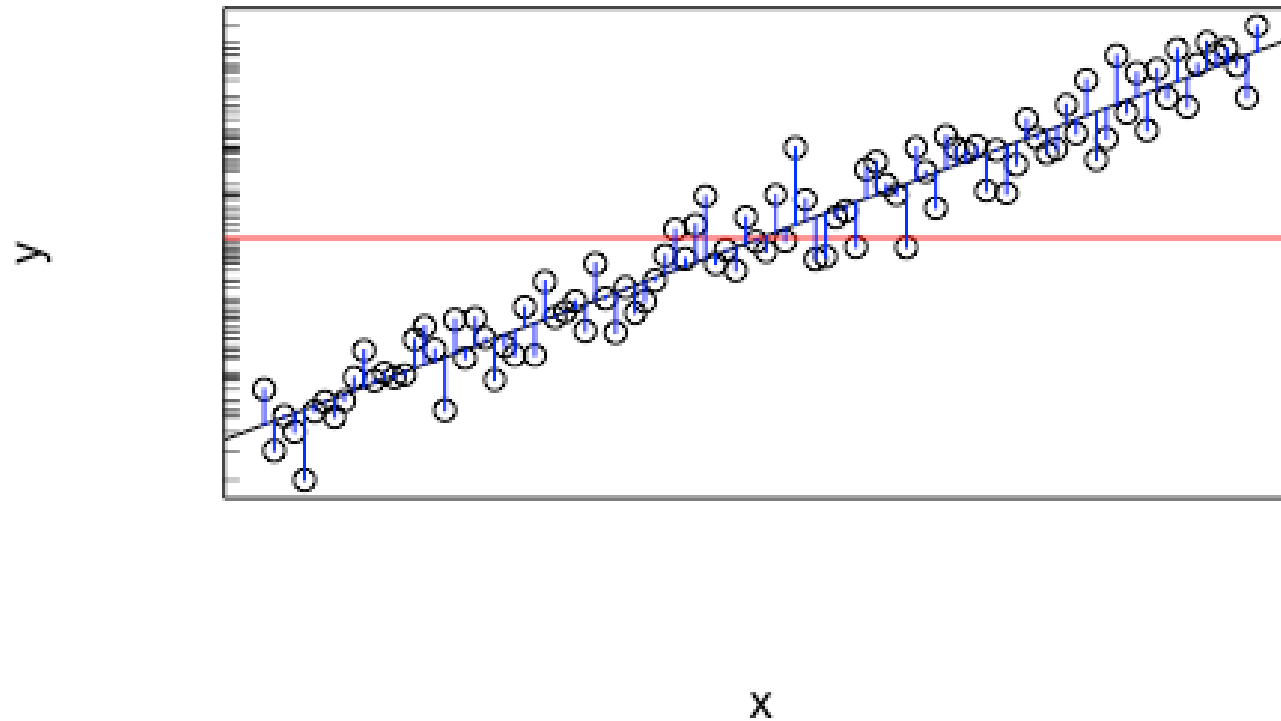
$$Precision = \frac{TP}{TP + FP}$$

Полнота показывает, какая часть положительных объектов была выделена классификатором:

$$Recall = \frac{TP}{TP + FN}$$

Метрика регрессии

MSE vs Var(Y)



Валидация

