



ПРОГРАММИРОВАНИЕ CUDA
C/C++, АНАЛИЗ ИЗОБРАЖЕНИЙ
И DEEP LEARNING

Лекция №6

Спасёнов Алексей

Метрики

1. Классификация
2. Регрессия
3. Кластеризация

При выборе алгоритма необходимо знать, какую метрику она оптимизирует

Классификация

1. Accuracy
2. Precision
3. Recall
4. ROC-AUC
5. F-метрика
6. Logloss

Регрессия

1. MAE
2. MSE
3. RMSE
4. RMSLE
5. R2

Классификация

Ассурасу

Подсчитываем долю правильно предсказанных объектов

Может быть использована в многоклассовой классификации

```
1. import numpy as np
2. target = np.array([1, 3, 2, 2, 3, 4, 1, 2])
3. pred = np.array([1, 3, 1, 2, 3, 2, 1, 2])
4. print(np.equal(target, pred).sum())
5. print(np.equal(target, pred).sum()/float(target.shape[0]))
```

Результат:

6

0.75

Когда могут возникнуть проблемы?

Классификация

Ассурасу

```
1. import numpy as np
2. target = np.array([1, 1, 1, 2, 1, 1, 1, 2])
3. pred = np.array([1, 1, 1, 2, 1, 1, 1, 1])
4. print(np.equal(target, pred).sum())
5. print(np.equal(target, pred).sum()/float(target.shape[0]))
```

Результат:

7

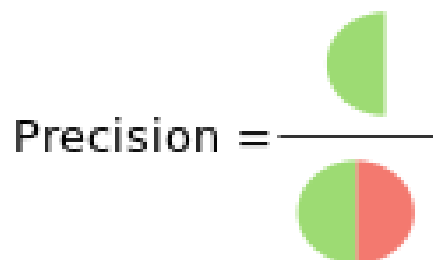
0.875

Классификация

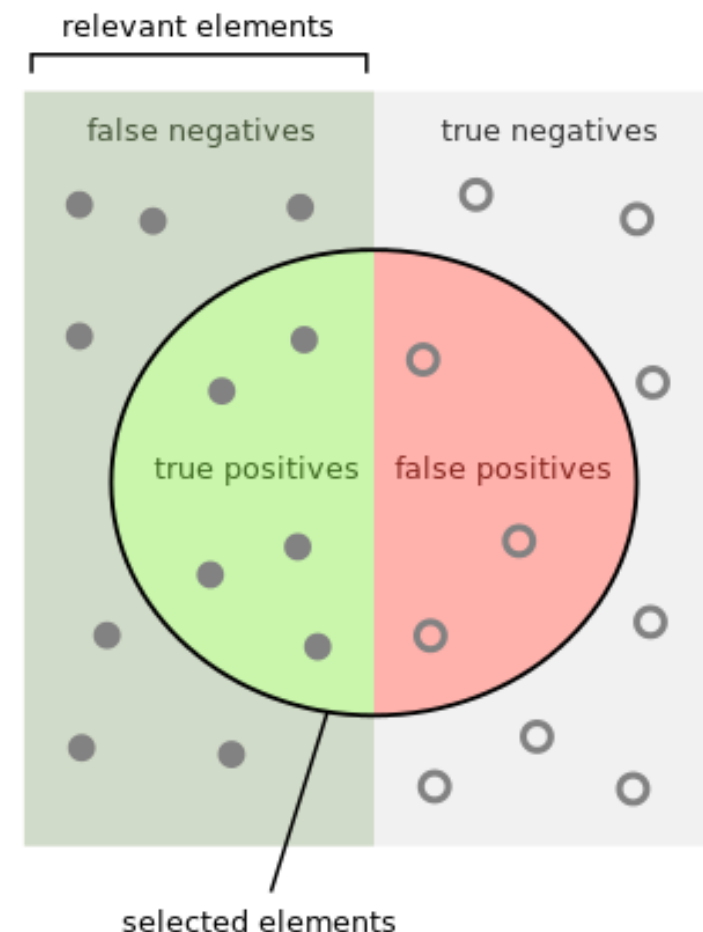
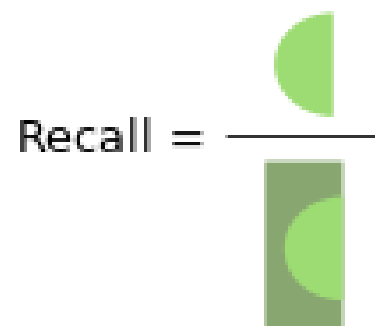
Precision, Recall

Задачи бинарной классификации

How many selected items are relevant?



How many relevant items are selected?



Классификация

Precision, Recall

	Предсказали True	Предсказали False
Ожидали True	True Positive (<i>tp</i>)	False Negative (<i>fn</i>)
Ожидали False	False Positive (<i>fp</i>)	True Negative (<i>tn</i>)

Полнота $Recall = \frac{tp}{tp + fn}$

Recall: Какую часть из объектов класса 1 мы нашли?

Точность $Precision = \frac{tp}{tp + fp}$

Precision: Какая часть из тех объектов класса 1, которую мы нашли, действительно принадлежат этому классу?

Классификация

Precision, Recall

```
1. target = np.array([0, 1, 1, 0, 1, 1])
2. pred = np.array([1, 0, 1, 0, 1, 0])
3. print(recall(target, pred))
4. print(precision(target, pred))
```

Результат:

0.5 $2/(2+2)$

0.67 $2/(2+1)$

Recall: Какую часть из объектов класса 1 мы нашли?

Precision: Какая часть из тех объектов класса 1, которую мы нашли, действительно принадлежат этому классу?

Классификация

Precision, Recall

```
1. target = np.array([0, 1, 1, 0, 1, 1])
2. pred = np.array([1, 1, 1, 1, 1, 1])
3. print(recall(target, pred))
4. print(precision(target, pred))
```

Результат:

0.1 $4/(4+0)$

0.67 $4/(4+2)$

Recall: Какую часть из объектов класса 1 мы нашли?

Precision: Какая часть из тех объектов класса 1, которую мы нашли, действительно принадлежат этому классу?

Классификация

$F1$ -score

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

```
1. target = np.array([0, 1, 1, 0, 1, 1])
2. pred = np.array([1, 0, 1, 0, 1, 0])
3. print(f1_score(target, pred))
```

Результат:
0.57

```
1. target = np.array([0, 1, 1, 0, 1, 1])
2. pred = np.array([1, 1, 1, 1, 1, 1])
3. print(f1_score(target, pred))
```

Результат:
0.8

Классификация

$F1$ -score

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Skewness...

```
1. target = np.array([0, 1, 0, 0, 0, 0])
2. pred = np.array([1, 1, 1, 0, 1, 0])
3. print(f1_score(target, pred))
```

Результат:
0.4

```
1. target = np.array([0, 1, 0, 0, 0, 0])
2. pred = np.array([1, 1, 1, 1, 1, 1])
3. print(f1_score(target, pred))
```

Результат:
0.29

Классификация

F_β - score

$$F_\beta = (1 + \beta^2) * \frac{Precision * Recall}{(\beta^2)Precision + Recall}$$

$0 < \beta < 1$ – предпочтение точности (*Precision*)

$1 < \beta$ – предпочтение полноте (*Recall*)

Классификация

ROC – AUC

ROC - Receiver Operating Characteristic

AUC - Area Under the Curve

Определяет долю правильно отранжированных пар

$$TPR(Recall) = \frac{tp}{tp + fn}$$

The Relationship Between Precision-Recall and ROC Curves
Jesse Davis, Mark Goadrich

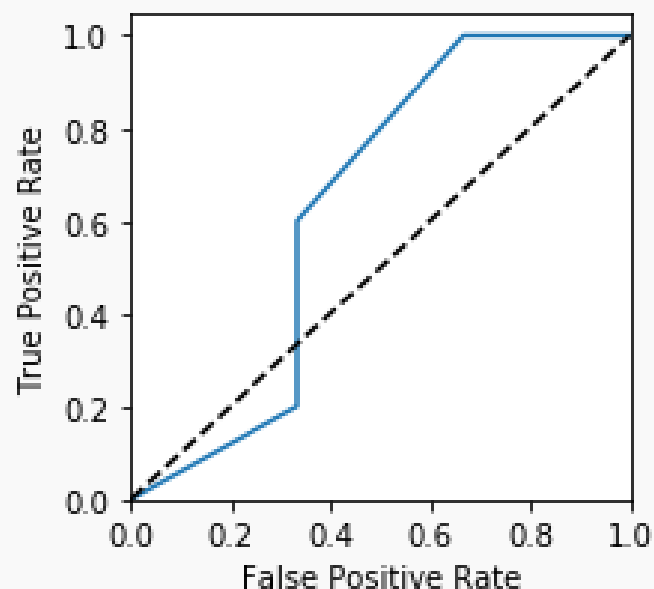
$$FPR = \frac{fp}{fp + tn}$$

<https://alexanderdyakonov.wordpress.com/2015/10/09/задачи-про-auc-roc/>

Классификация

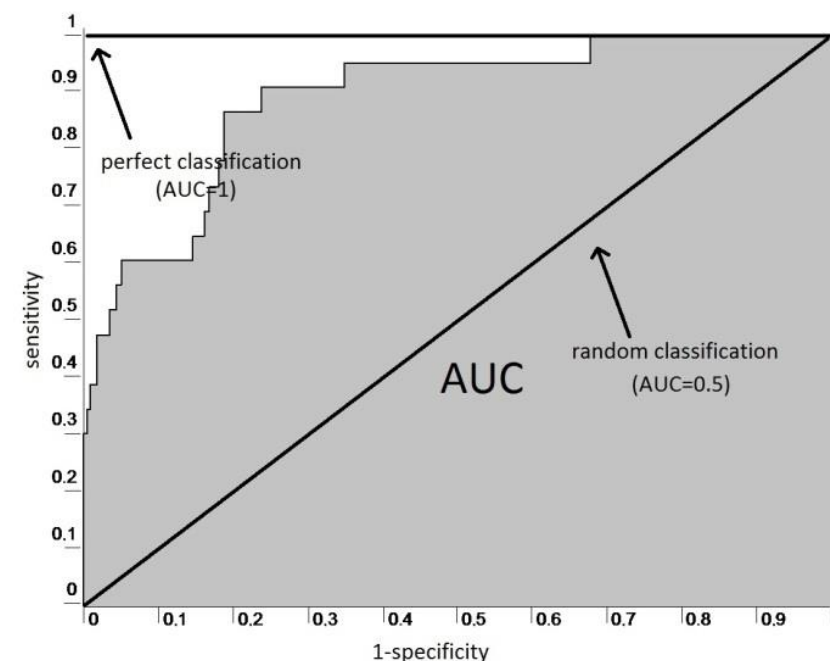
ROC – AUC

1. target = [0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0]
2. pred = [0.1, 0.3, 0.2, 0.4, 0.7, 0.8, 0.7, 0.8, 0.3, 0.4, 0.8]



$$TPR(Recall) = \frac{tp}{tp + fn}$$

$$FPR = \frac{fp}{fp + tn}$$



Классификация

Logloss Чем сильнее ошибаемся, тем больше ошибка

$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij})$$

N — размер выборки, M — количество классов

$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N (y_i \log(p_i) + (1 - y_i) \log(1 - p_i))$$

<https://www.kaggle.com/wiki/LogarithmicLoss>

Регрессия

MAE

Mean Absolute Error

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

Лучшее константное предсказание - медиана

\hat{Y}	Y
0.1	0
0.5	1
0.6	1
0.5	1
0.3	0

$$MAE = \frac{1}{5} (|0.1 - 0| + |0.5 - 1| + |0.6 - 1| + |0.5 - 1| + |0.3 - 0|)$$

Регрессия

MSE (Mean Squared Error)

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

RMSE (Root Mean Squared Error)

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

Лучшее константное предсказание - среднее

Введение в Машинное обучение



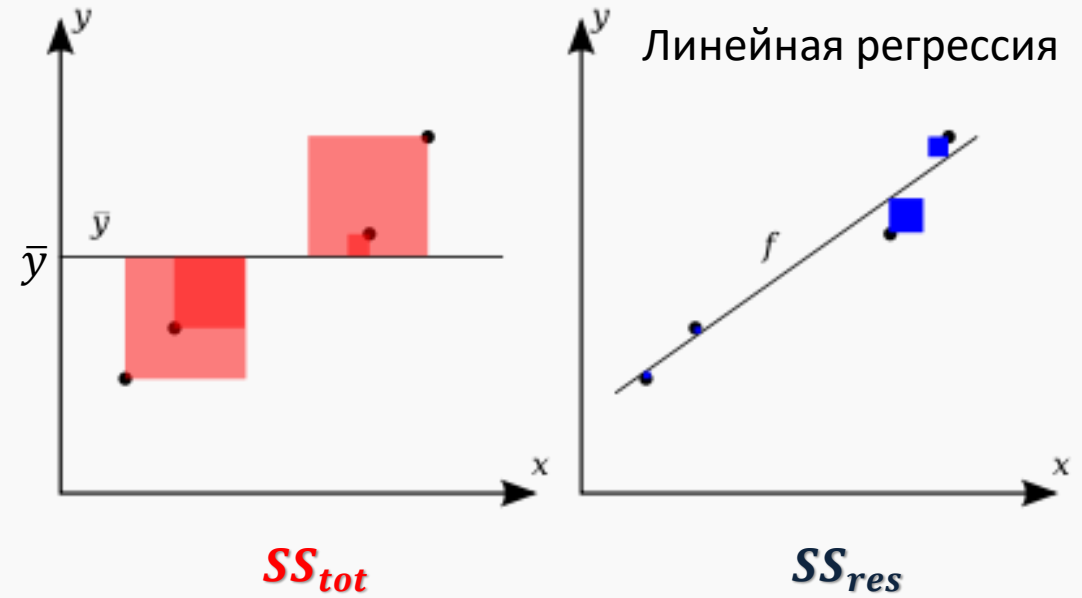
Регрессия

R2 – score

$$\bar{y} = \sum_{i=1}^N y_i$$

$$SS_{tot} = \sum_{i=1}^N (y_i - \bar{y})^2 - \text{variance}$$

$$SS_{res} = \sum_{i=1}^N (y_i - f_i)^2$$



$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$



Контакты:
a.spasenov@corp.mail.ru
[alex_spasenov](#) (Skype)



Спасибо за внимание!