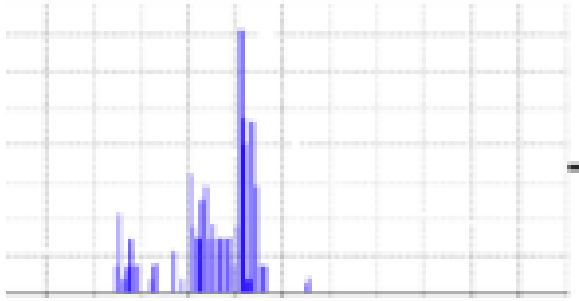
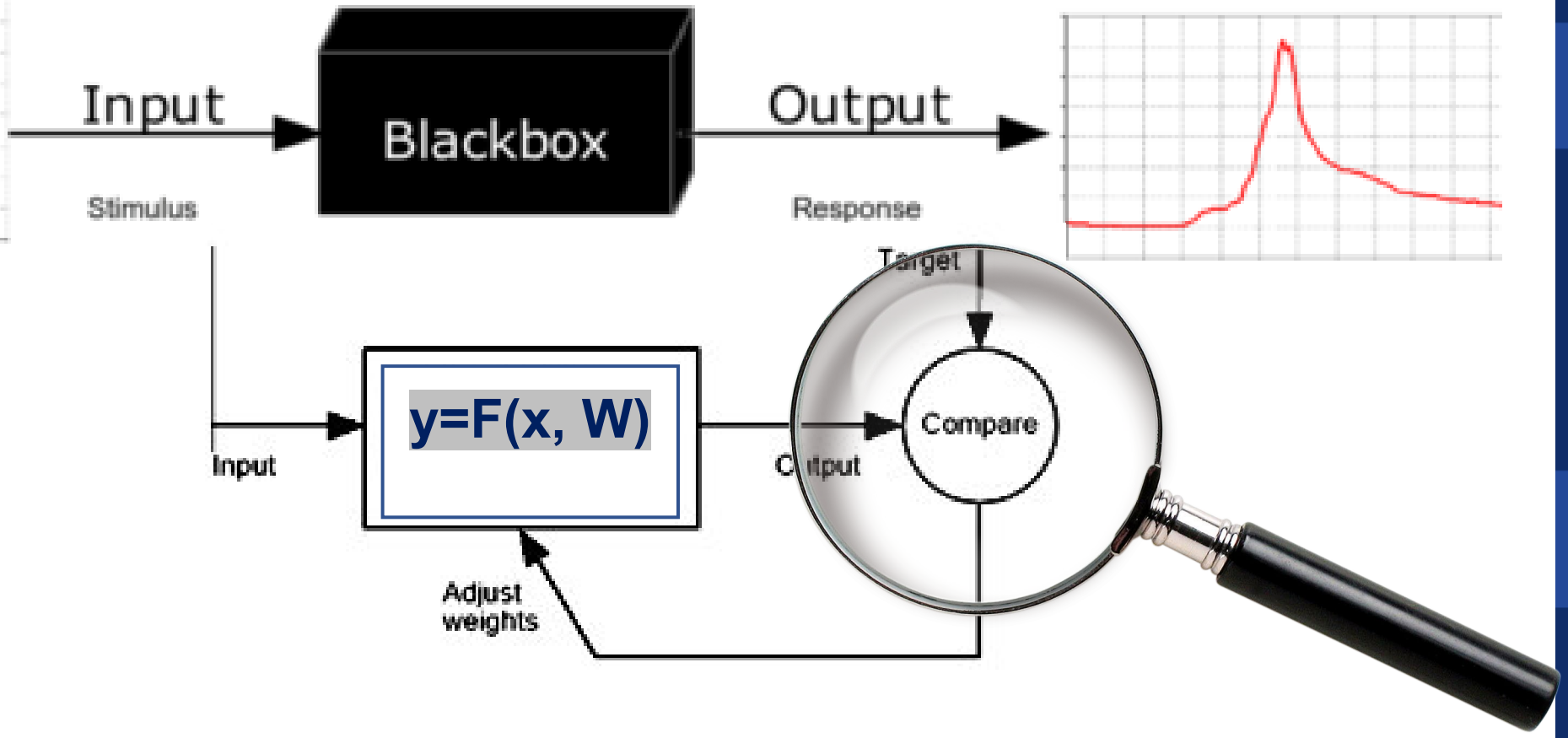
Лекция 04
Метрики и экспериментальный дизайн

А. Задачи
Б. Метрики регрессии
В. Метрики классификации
Г. Метрики кластеризации
Д. Метрики производительности

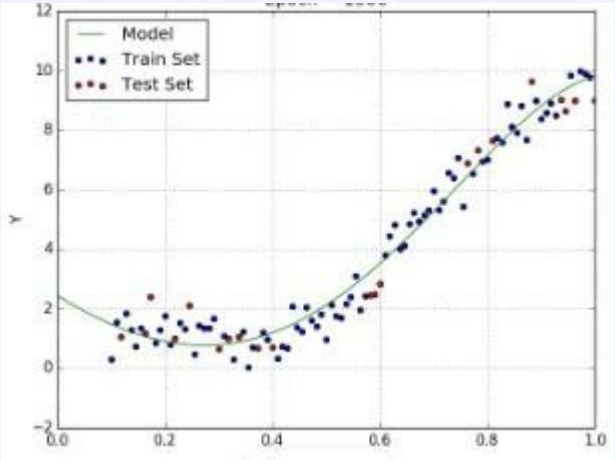
«Черный ящик»



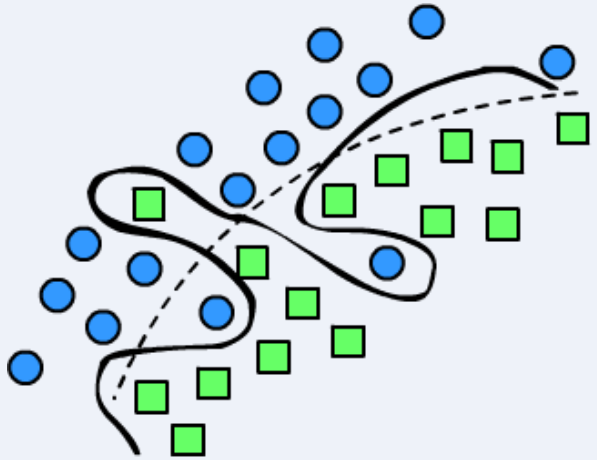

Input
Blackbox
Output
Stimulus
Response
Target
y=F(x, W)
Input
Compare
Adjust
weights

Задачи

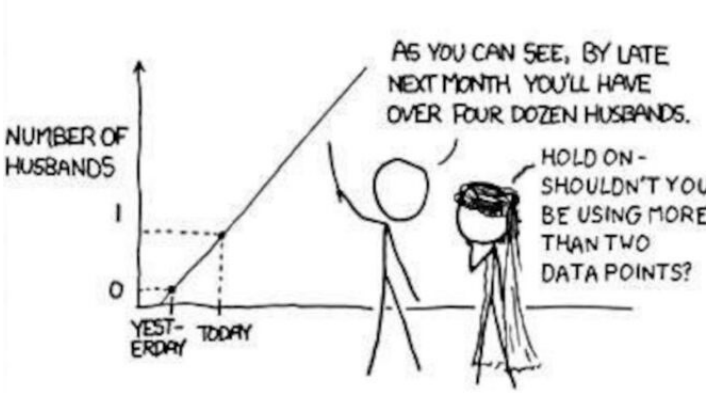
Регрессия:
непрерывные значения



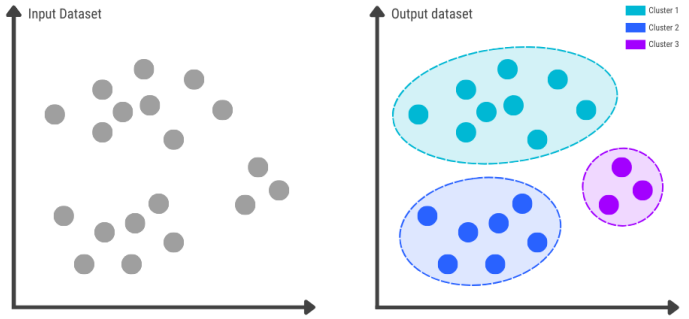
Классификация:
дискретные значения



Прогнозирование:
будущие значения



Кластеризация:
близость значений

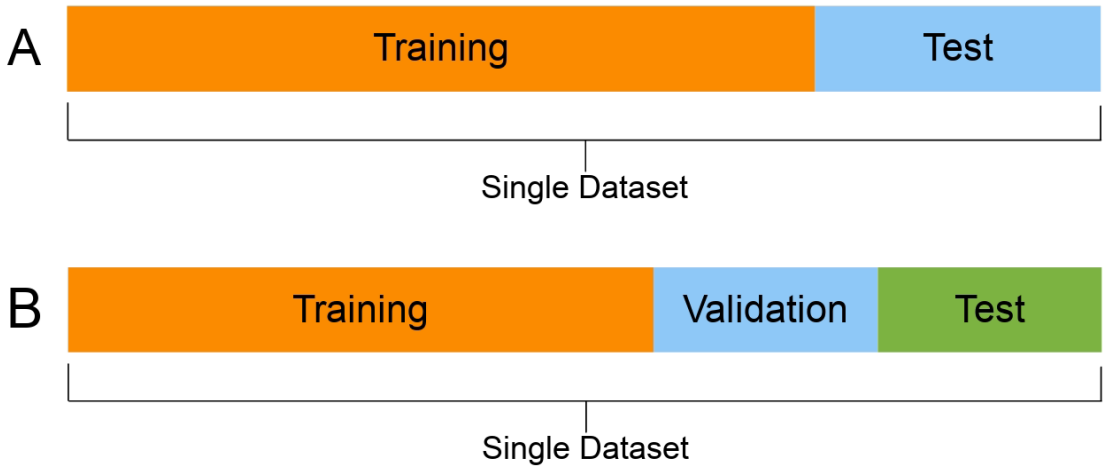


На одном – мгновенная
На наборе – усредненная
На последовательности - временная

Ранжирование:
порядок значений

true relevance	1	0	0	0	0
predicted relevance	0.7	0.5	0.5	0.5	0.5

Обучение VS тест. Метрики VS Функции ошибки



Функция ошибки Loss – чему учим

≠

Метрика metric – что наблюдаем

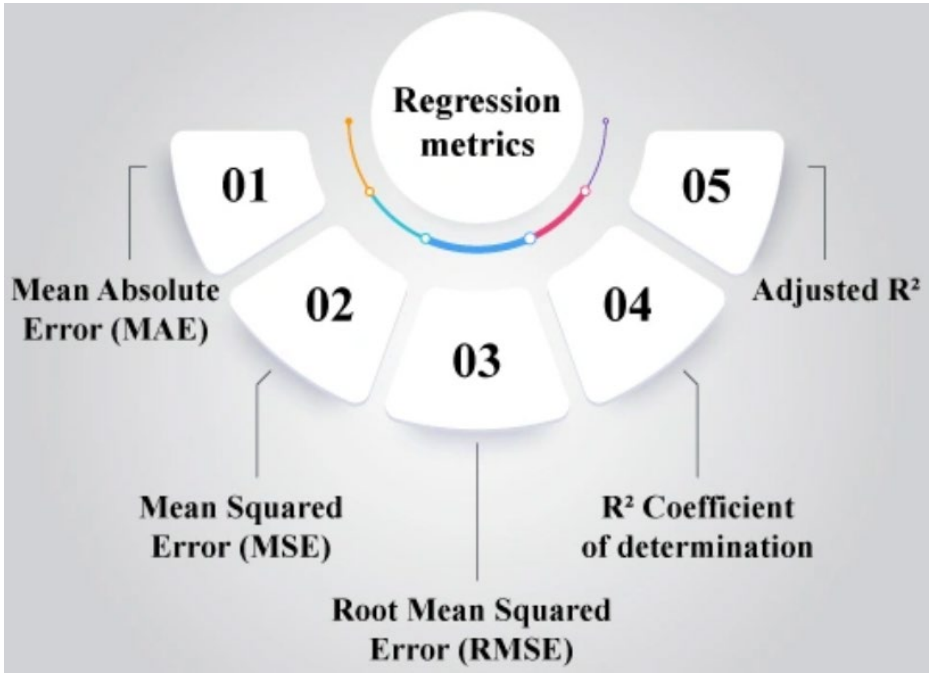
Метрики:

Прикладные – что действительно хотим

≠

Математические – что можем посчитать сейчас

Метрики регрессии



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

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

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

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

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \cdot 100\%$$

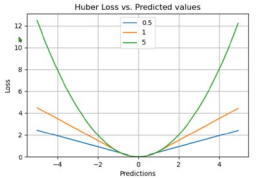
where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Normalized ….

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

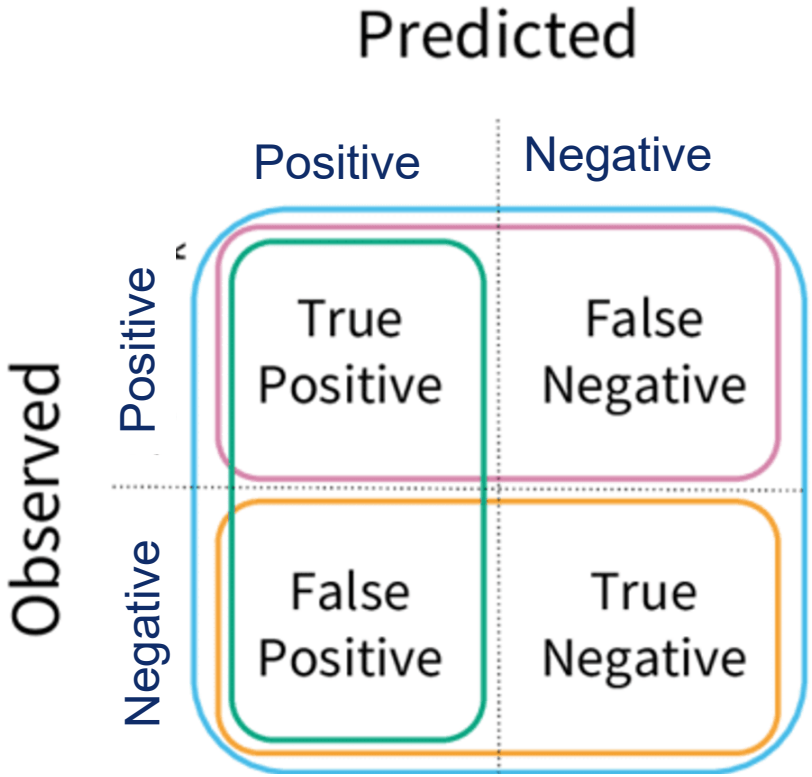
Huber Loss

$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{for} |y - f(x)| \le \delta, \\ \delta|y - f(x)| - \frac{1}{2}\delta^2 & \text{otherwise.} \end{cases}$$



https://www.analyticsvidhya.com/blog/2021/10/evaluation-metric-for-regression-models/

Метрики классификации. Бинарная



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

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

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

$$F\beta\ \text{Score} = \frac{(1 + \beta^2) \times \text{Precision} \times \text{Recall}}{(\beta^2 \times \text{Precision}) + \text{Recall}}$$

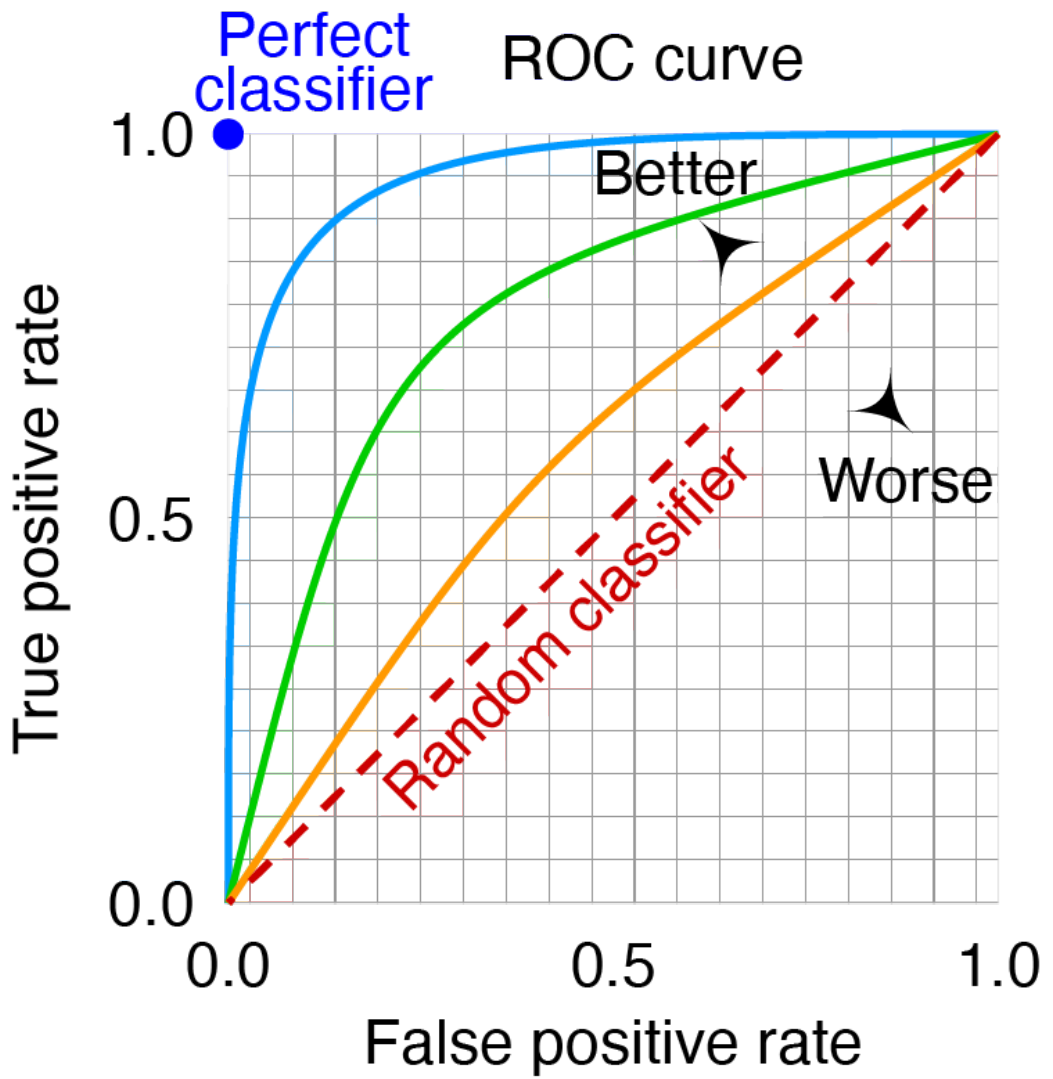
Метрики классификации. AUC

Color
Red
Red
Yellow
Green
Yellow

Red	Yellow	Green
1	0	0
1	0	0
0	1	0
0	0	1
0	1	0

LogLoss

$$-\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$



https://towardsdatascience.com/understanding-auc-scores-in-depth-whats-the-point-5f2505eb499f/

Метрики классификации. Многоклассовая

	precision	recall	f1-score	support
Aeroplane	0.67	0.67	0.67	3
Boat	0.25	1.00	0.40	1
Car	1.00	0.50	0.67	6
accuracy			0.60	10
macro avg	0.64	0.72	0.58	10
weighted avg	0.82	0.60	0.64	10

Average F1 scores

One-vs-Rest (OvR)

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Precision	Recall	F1 Score
Airplane	2	1	1	0.67	0.67	2 * (0.67 * 0.67) / (0.67 + 0.67) = **0.67**
Boat	1	3	0	0.25	1.00	2*(0.25 * 1.00) / (0.25 + 1.00) = **0.40**
Car	3	0	3	1.00	0.50	2 * (1.00 * 0.50) / (1.00 + 0.50) = **0.67**

		Predicted: Airplane	Predicted: Boat	Predicted: Car
Actual	Airplane	2	1	0
	Boat	0	1	0
	Car	1	2	3

Label	Per-Class F1 Score	Macro-Averaged F1 Score
Airplane	0.67	$\frac{0.67 + 0.40 + 0.67}{3}$ = **0.58**
Boat	0.40	
Car	0.67	

Label	Per-Class F1 Score	Support	Support Proportion	Weighted Average F1 Score
Airplane	0.67	3	0.3	$(0.67 * 0.3) + (0.40 * 0.1) + (0.67 * 0.6)$ = **0.64**
Boat	0.40	1	0.1	
Car	0.67	6	0.6	
Total	-	10	1.0	

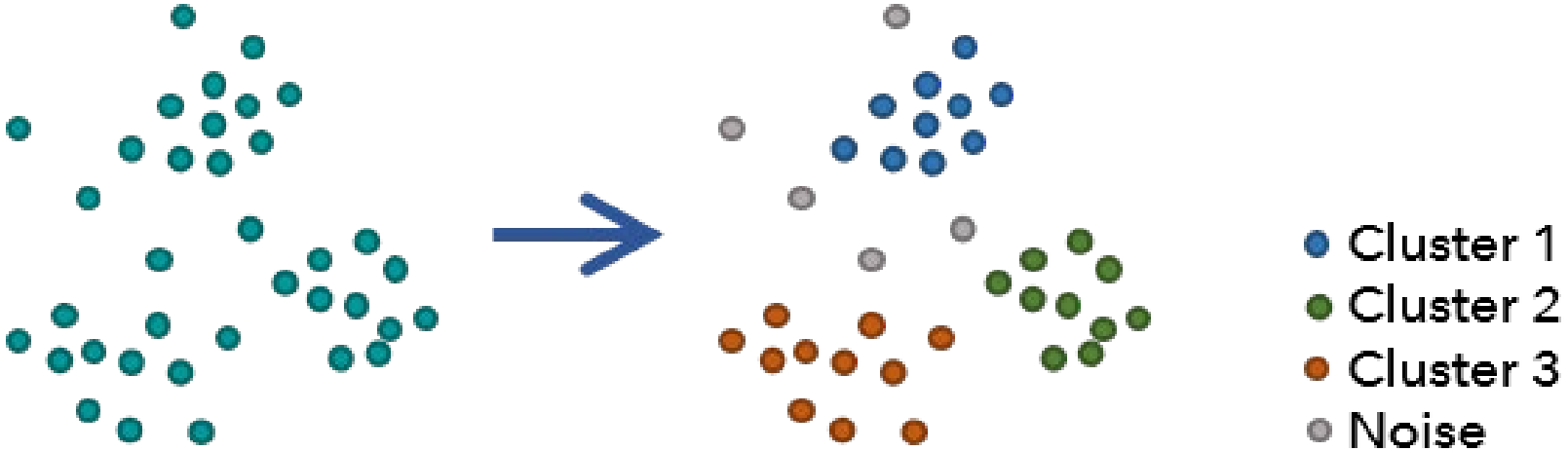
Label	True Positive (TP)	False Positive (FP)	False Negative (FN)	Micro-Averaged F1 Score
Airplane	2	1	1	$\frac{TP}{TP+\frac{1}{2}(FP+FN)} = \frac{6}{6+\frac{1}{2}(4+4)}$ = **0.60**
Boat	1	3	0	
Car	3	0	3	
TOTAL	**6**	**4**	**4**	

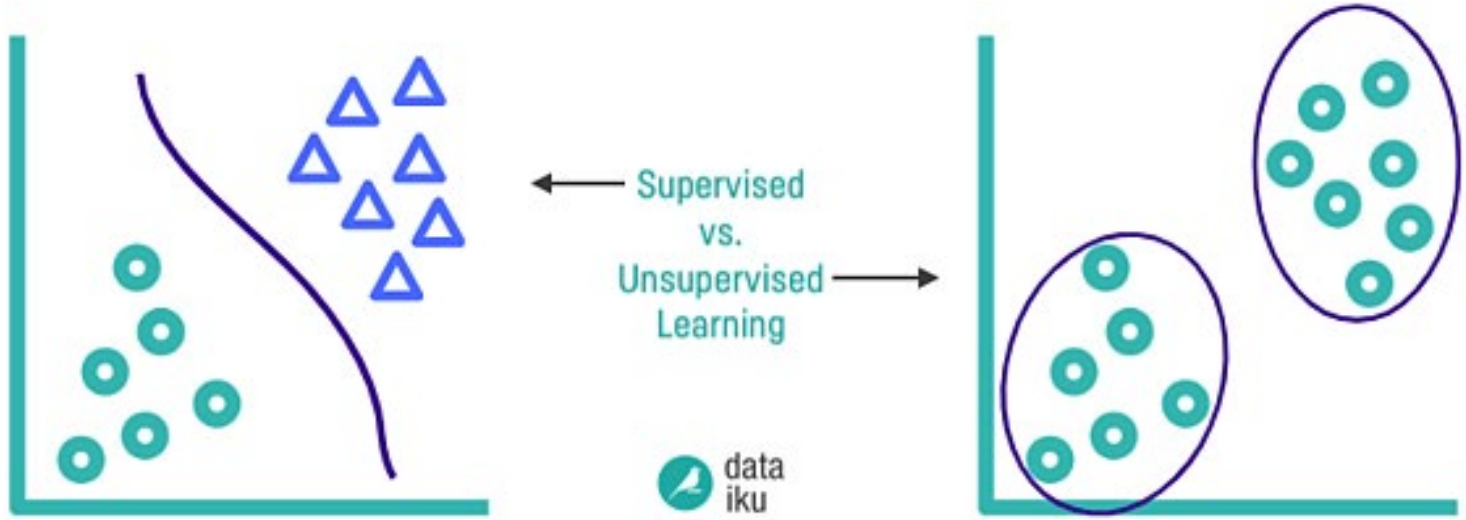
$$\text{F1 Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

Не надо спрашивать «Какая лучше?». Публикуйте все.

https://towardsdatascience.com/micro-macro-weighted-averages-of-f1-score-clearly-explained-b603420b292f/

Метрики кластеризации

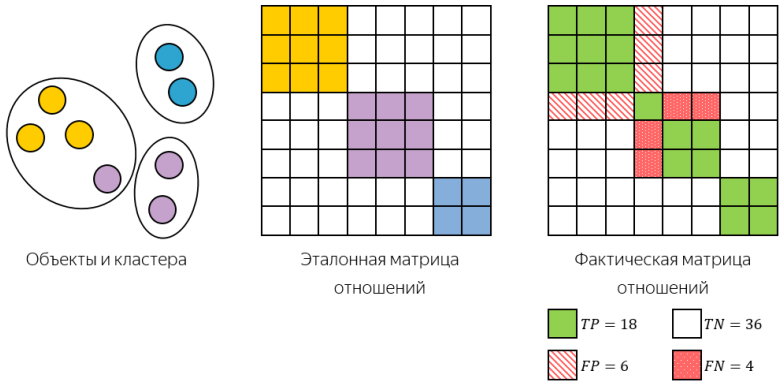




Метрики кластеризации

ВНЕШНИЕ

Есть эталонные кластеры (классы) (редкость!)



$X \backslash Y$	Y_1	Y_2	...	Y_s	Sums
X_1	n_{11}	n_{12}	...	n_{1s}	a_1
X_2	n_{21}	n_{22}	...	n_{2s}	a_2
⋮	⋮	⋮	⋱	⋮	⋮
X_r	n_{r1}	n_{r2}	...	n_{rs}	a_r
Sums	b_1	b_2	...	b_s	n

$$Rand = \frac{TP + TN}{TP + TN + FP + FN} \qquad Jaccard = \frac{TP}{TP + FN + FP}$$

Пусть $p_{ij} = \frac{n_{ij}}{n}, p_i = \frac{a_i}{n}, p_j = \frac{b_j}{n}$.

- Элементы принадлежат одному кластеру и одному классу — TP
- Элементы принадлежат одному кластеру, но разным классам — FP
- Элементы принадлежат разным кластерам, но одному классу — FN
- Элементы принадлежат разным кластерам и разным классам — TN

$$\overbrace{ARI}^{\text{Adjusted Index}} = \frac{\overbrace{\sum_{ij}\binom{n_{ij}}{2}}^{\text{Index}} - \overbrace{[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}]/\binom{n}{2}}^{\text{Expected Index}}}{\underbrace{\frac{1}{2}[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}]}_{\text{Max Index}} - \underbrace{[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}]/\binom{n}{2}}_{\text{Expected Index}}}$$

$$E = -\sum_i p_i \left(\sum_j \frac{p_{ij}}{p_i} log\left(\frac{p_{ij}}{p_i}\right)\right)$$

однородность

Пары примеров. Число перестановок: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

https://habr.com/ru/company/yandex/blog/500742/

https://neerc.ifmo.ru/wiki/index.php?title=Оценка_качества_в_задаче_кластеризации

Метрики кластеризации

ВНУТРЕННИЕ

Нет эталонных кластеров.

Сравниваем кластеры между собой и их центроиды

Компактность

$$WSS = \sum_{j=1}^{M} \sum_{i=1}^{|C_j|} (x_{ij} - \overline{x_j})^2$$

Отделимость

$$BSS = n \cdot \sum_{j=1}^{M} (\overline{x_j} - \overline{x})^2$$

Силуэт

a – среднее до точек внутри кластера
b – минимальное среднее до точек другого кластера

$$a(x_i, c_k) = \frac{1}{|c_k|} \sum_{x_j \in c_k} \|x_i - x_j\|$$

$$b(x_i, c_k) = min_{c_l \in C \setminus c_k} \{ \frac{1}{|c_l|} \sum_{x_j \in c_l} \|x_i - x_j\| \}$$

$$Sil(C) = \frac{1}{N} \sum_{c_k \in C} \sum_{x_i \in c_k} \frac{b(x_i, c_k) - a(x_i, c_k)}{max\{a(x_i, c_k), b(x_i, c_k)\}},$$

https://habr.com/ru/company/yandex/blog/500742/

Kullback Leibler Divergence Loss

$$\mathcal{L}_{KL}(x_m, x_n) = \sum_{j=1}^{C} \mathbf{s}_m^j * \log \frac{\mathbf{s}_m^j}{\mathbf{s}_n^j}.$$

Индекс Дэвиса-Болдуина

$$DB(C) = \frac{1}{K} \sum_{c_k \in C} \max_{c_l \in C \setminus c_k} \left\{ \frac{S(c_k) + S(c_l)}{\|\overline{c_k} - \overline{c_l}\|} \right\},$$

где:

$$S(c_k) = \frac{1}{|c_k|} \sum_{x_i \in c_k} \|x_i - \overline{c_k}\|$$

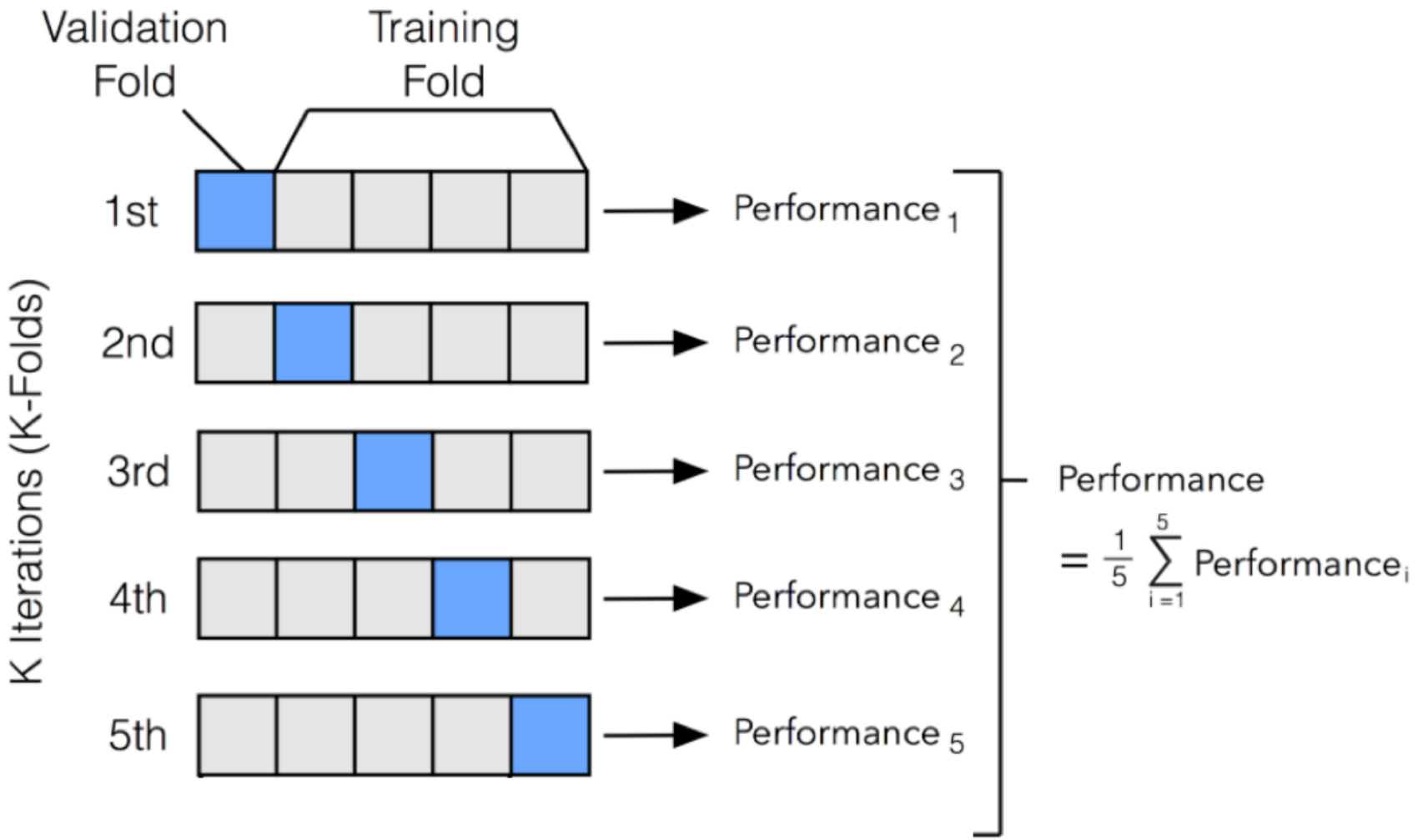
https://www.geeksforgeeks.org/machine-learning/clustering-metrics/

Метрики производительности

- Размер модели
- Размерность данных (входы, выходы, эмбединги и т.п.)
- Время обучения, частота переобучения и т.п.
- Время инференса (общее и латентность)
- Потребляемая память (по видам)
- ...

Хороший инженер протоколирует ВСЕ

Кроссвалидация



Q&A и контакты

Группа по дисциплине:

https://t.me/+8dShF1tFSDg0ZmJi