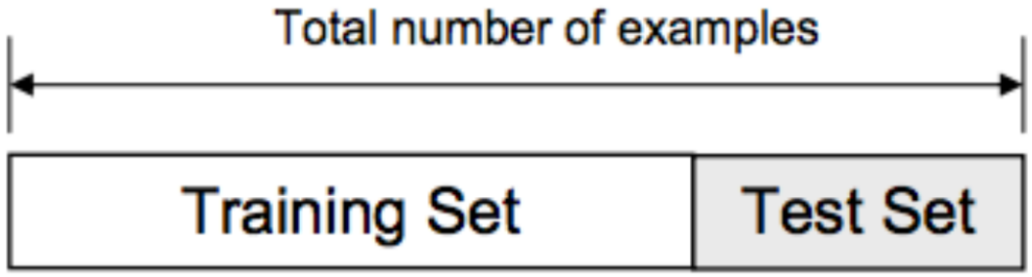
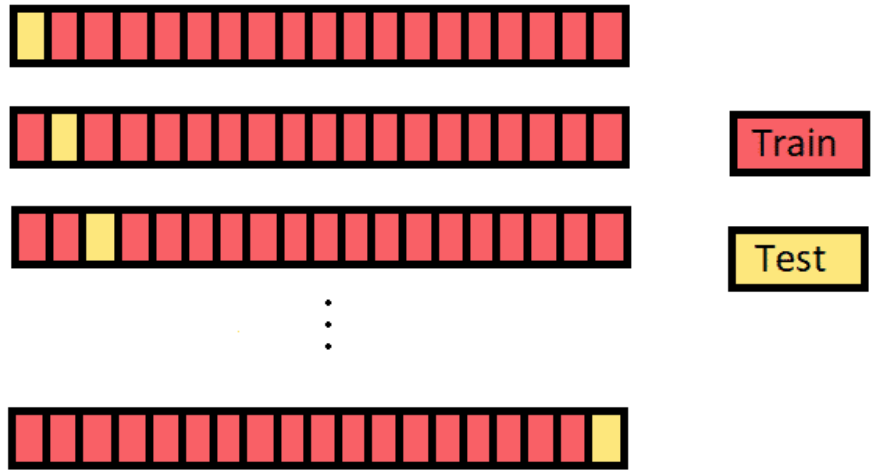
Метрики

Сбертех, МФТИ

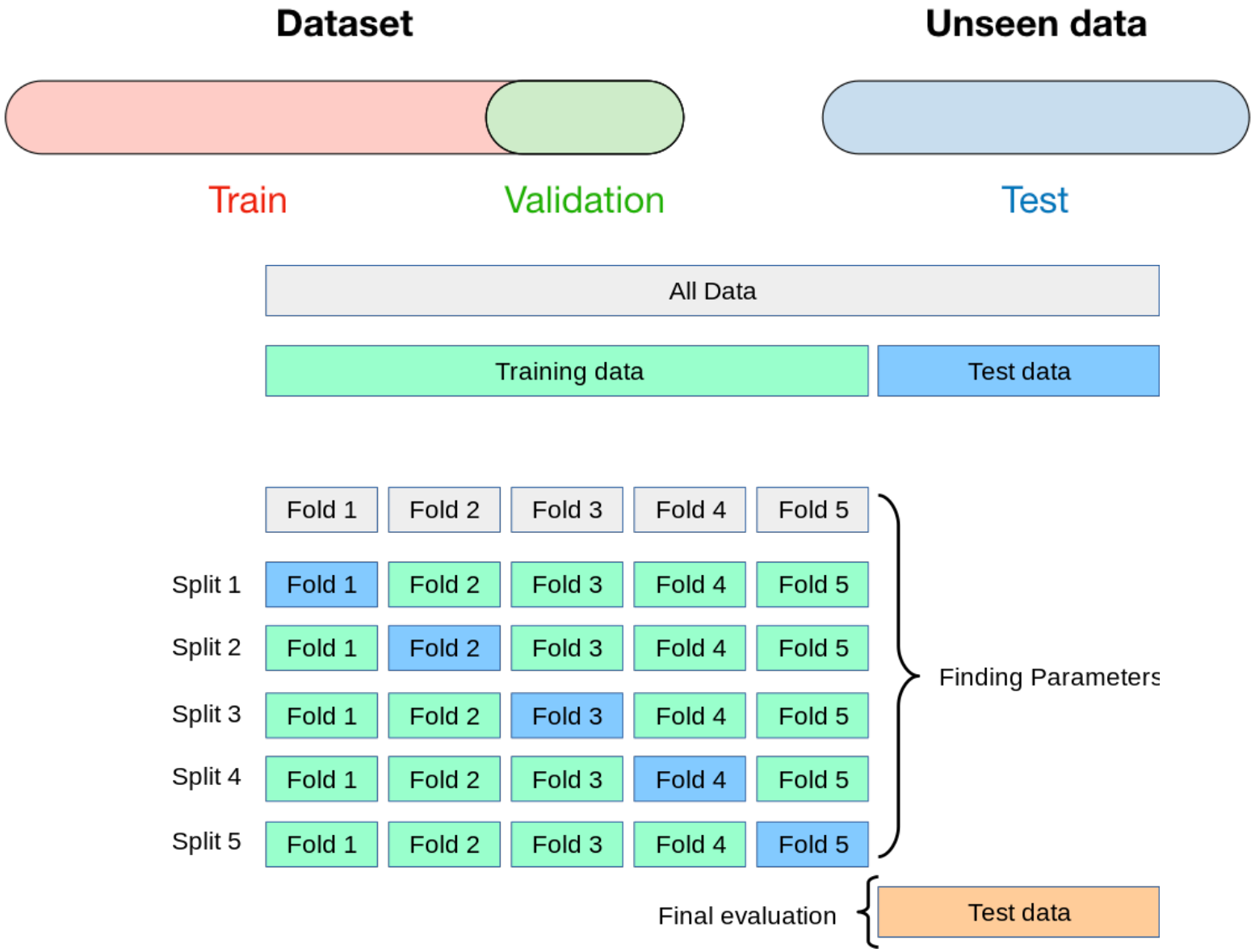
Стратегия валидации

- Train-test split
- Cross-validation
- Leave-one-out

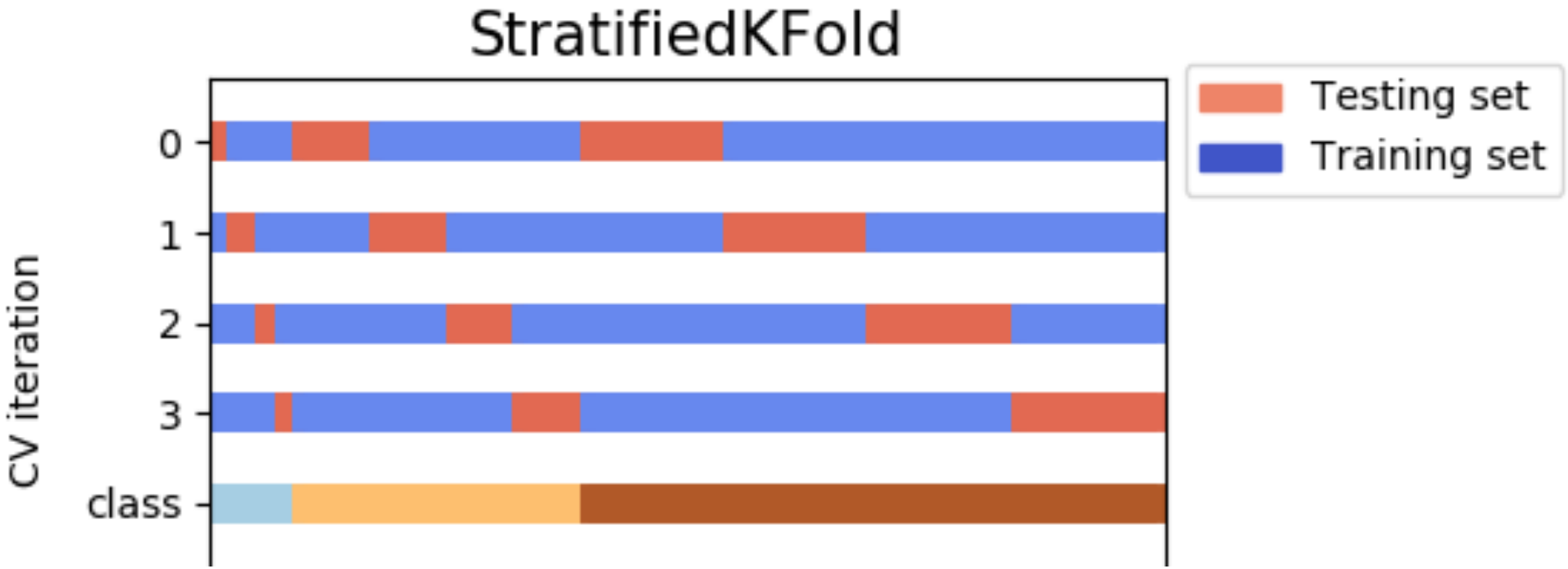




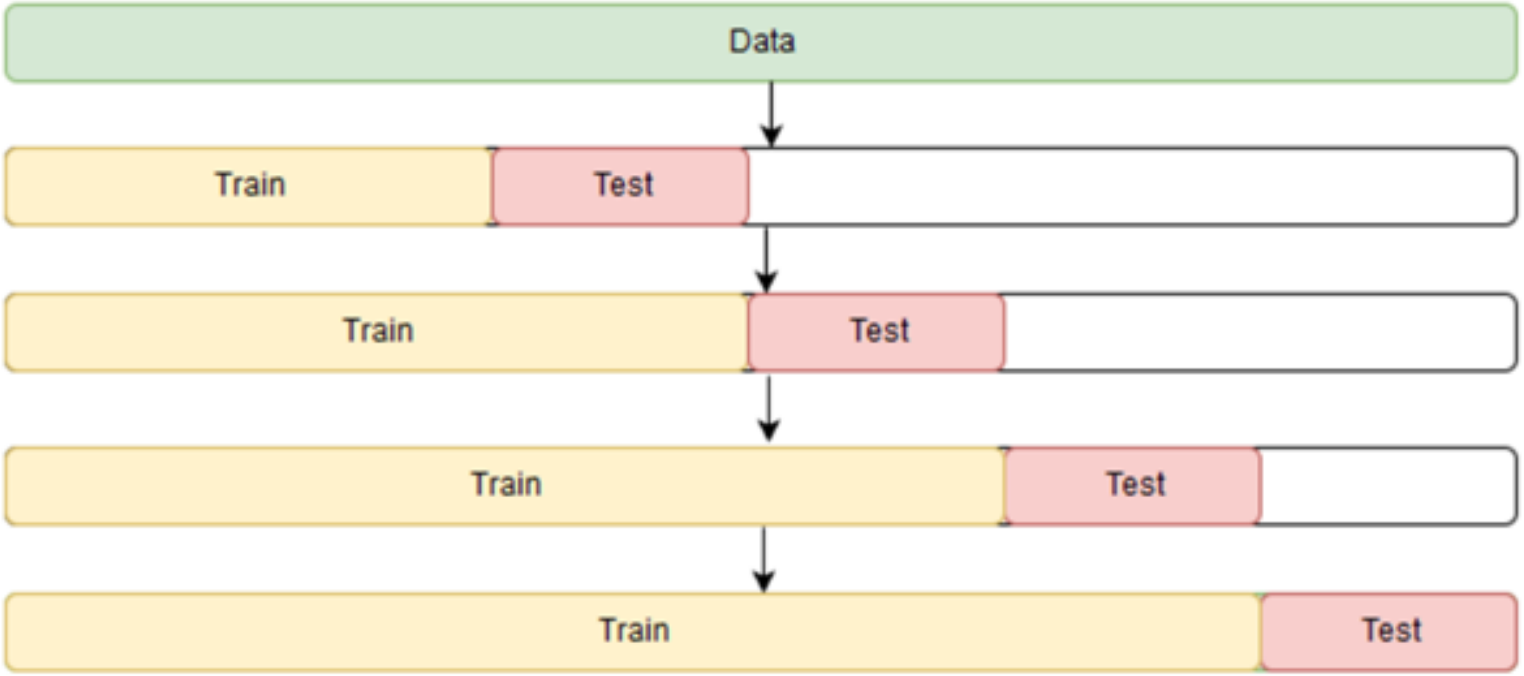
Стратегия валидации



Стратегия валидации

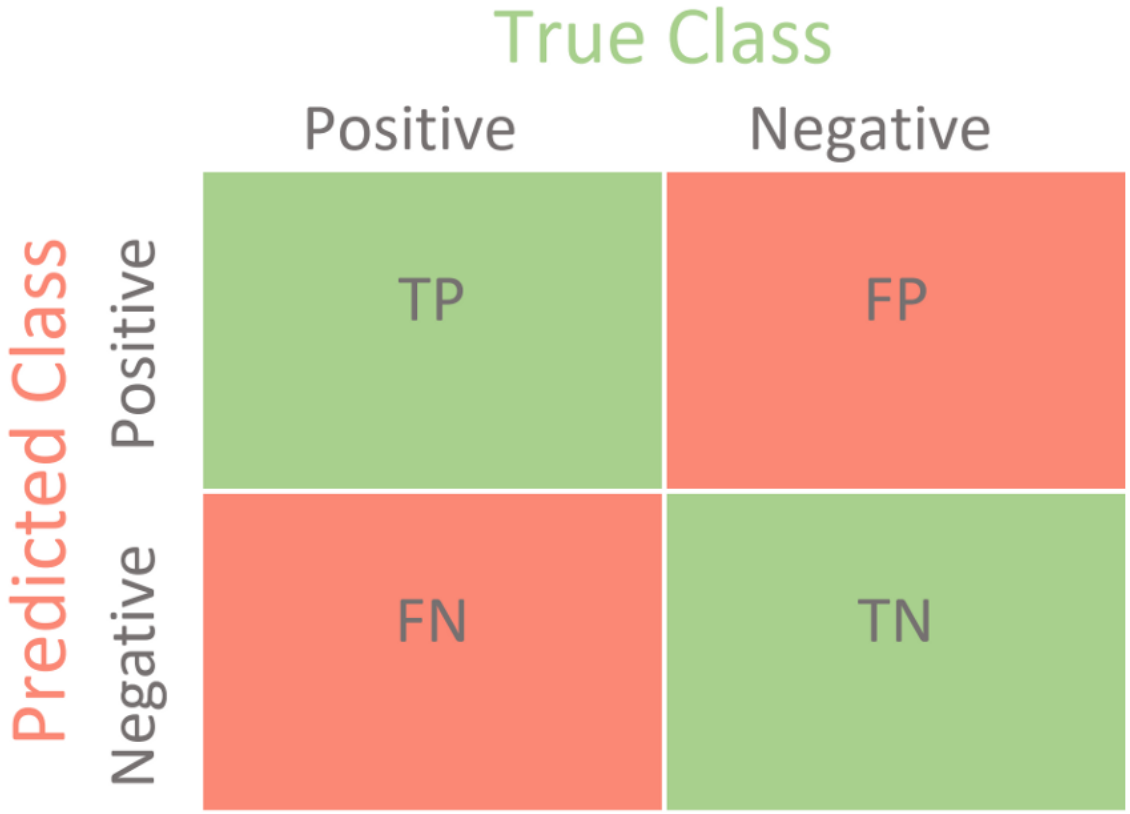


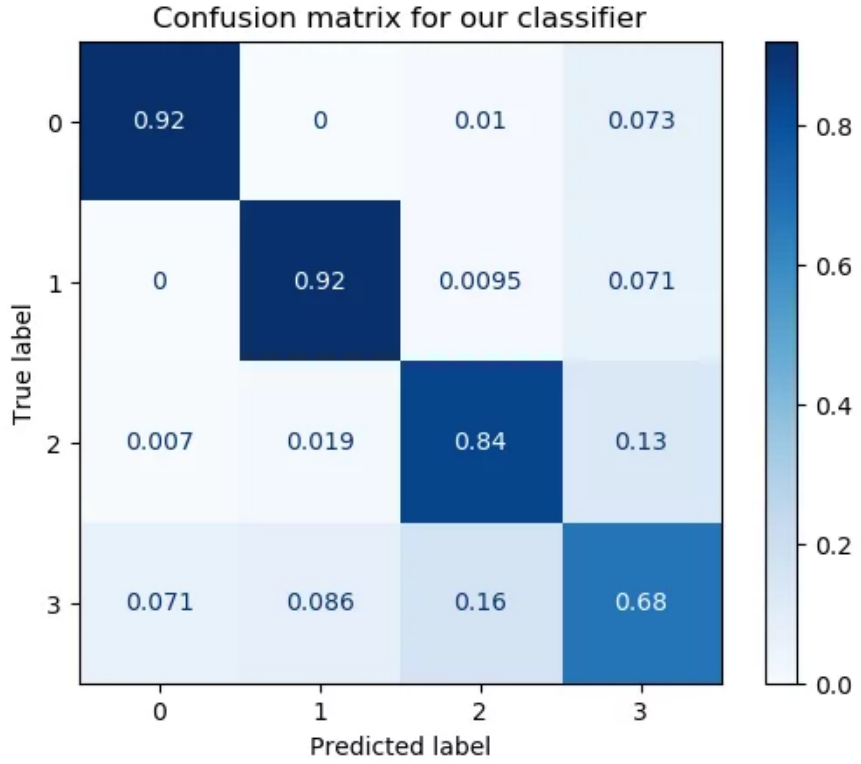
Time series



Метрики бинарной классификации

$$Accuracy(\hat{y}, y) = \frac{1}{\ell}\sum_{i=1}^{\ell}[\hat{y}_i = y_i]$$





Метрики бинарной классификации

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

$$F_\beta = (1 + \beta^2)\frac{precision.recall}{\beta^2.precision + recall}$$

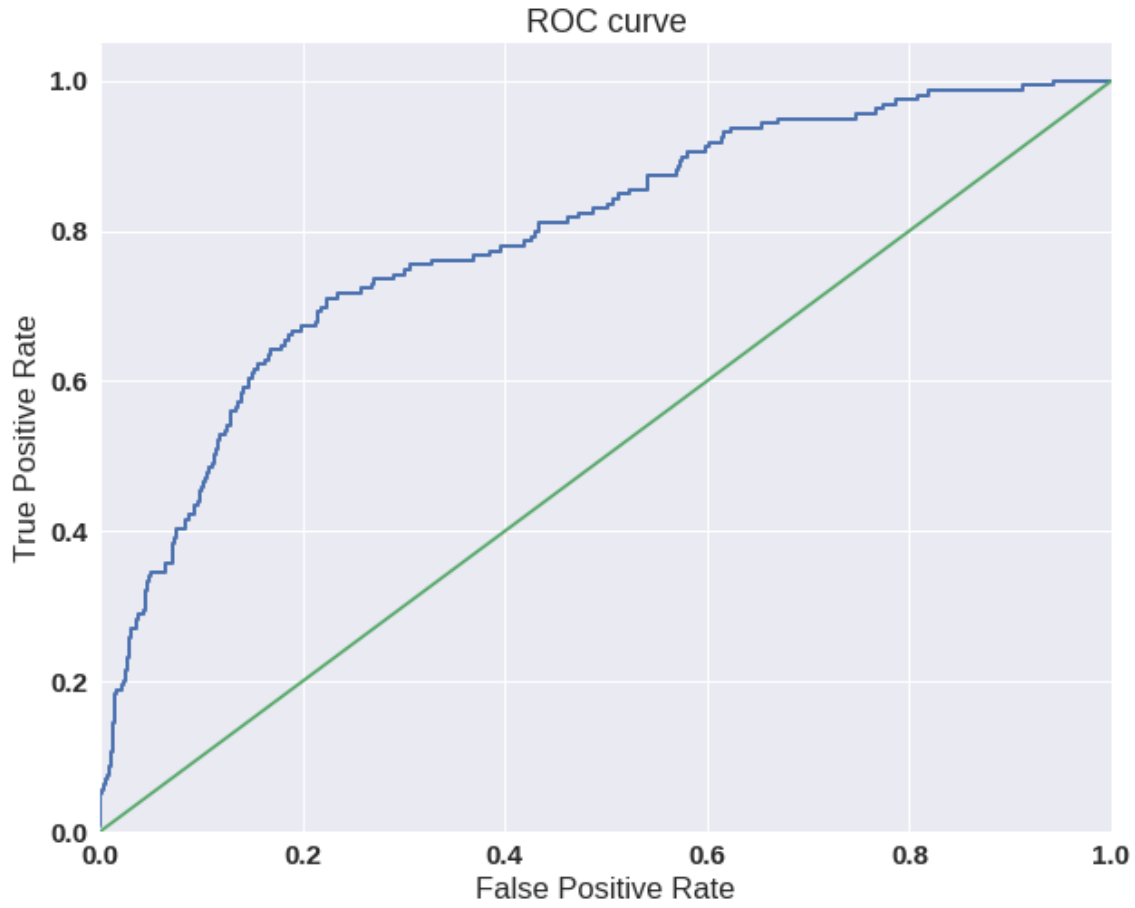
$$recall = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2.precision.recall}{precision + recall}$$

Метрики бинарной классификации

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

$$FPR = \frac{FP}{FP + TN}$$



https://alexanderdyakonov.wordpress.com/2017/07/28/auc-roc-площадь-под-кривой-ошибок/

Метрики бинарной классификации

$$TPR = \frac{TP}{TP + FN} = \frac{90}{90 + 10} = 0.9$$

$$precision = \frac{TP}{TP + FP} = 90/(90 + 10) = 0.9$$

$$FPR = \frac{FP}{FP + TN} = \frac{10}{10 + 999890} = 0.00001$$

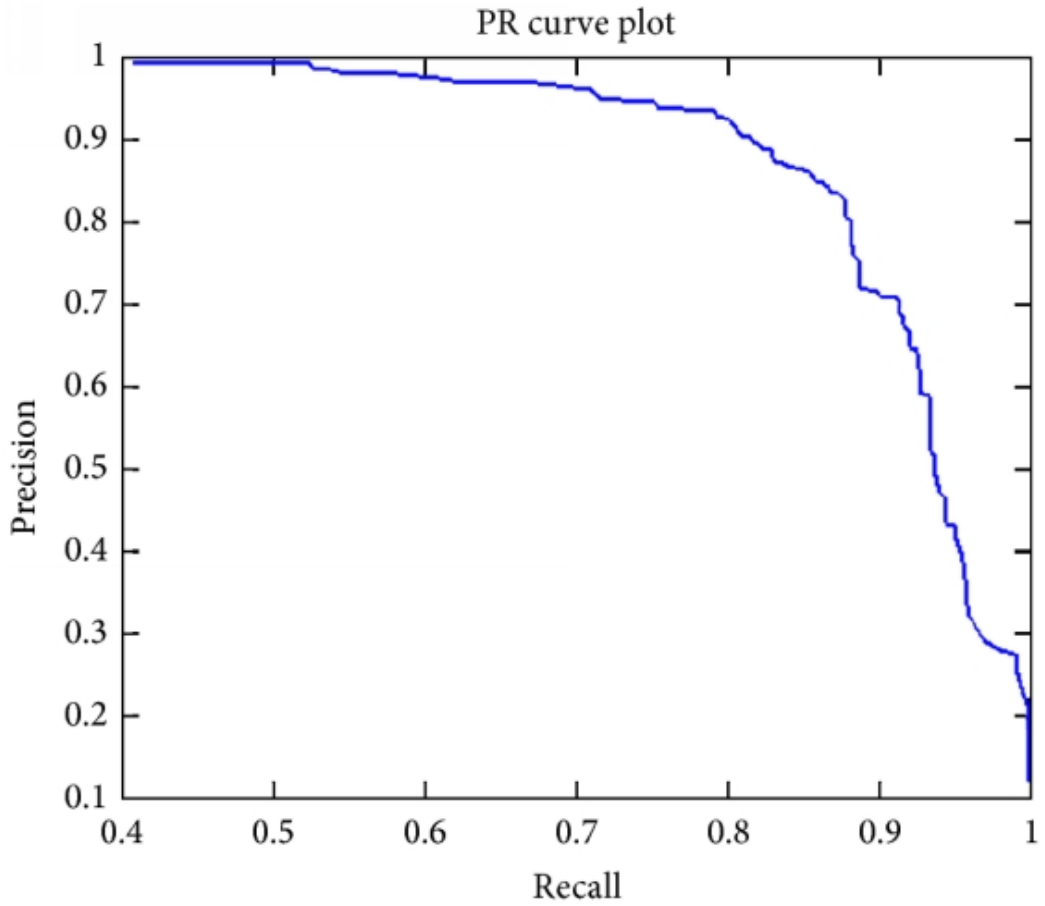
$$recall = \frac{TP}{TP + FN} = 90/(90 + 10) = 0.9$$

$$TPR = \frac{TP}{TP + FN} = \frac{90}{90 + 10} = 0.9$$

$$precision = \frac{TP}{TP + FP} = \frac{90}{90 + 1910} = 0.045$$

$$FPR = \frac{FP}{FP + TN} = \frac{1910}{1910 + 997990} = 0.001$$

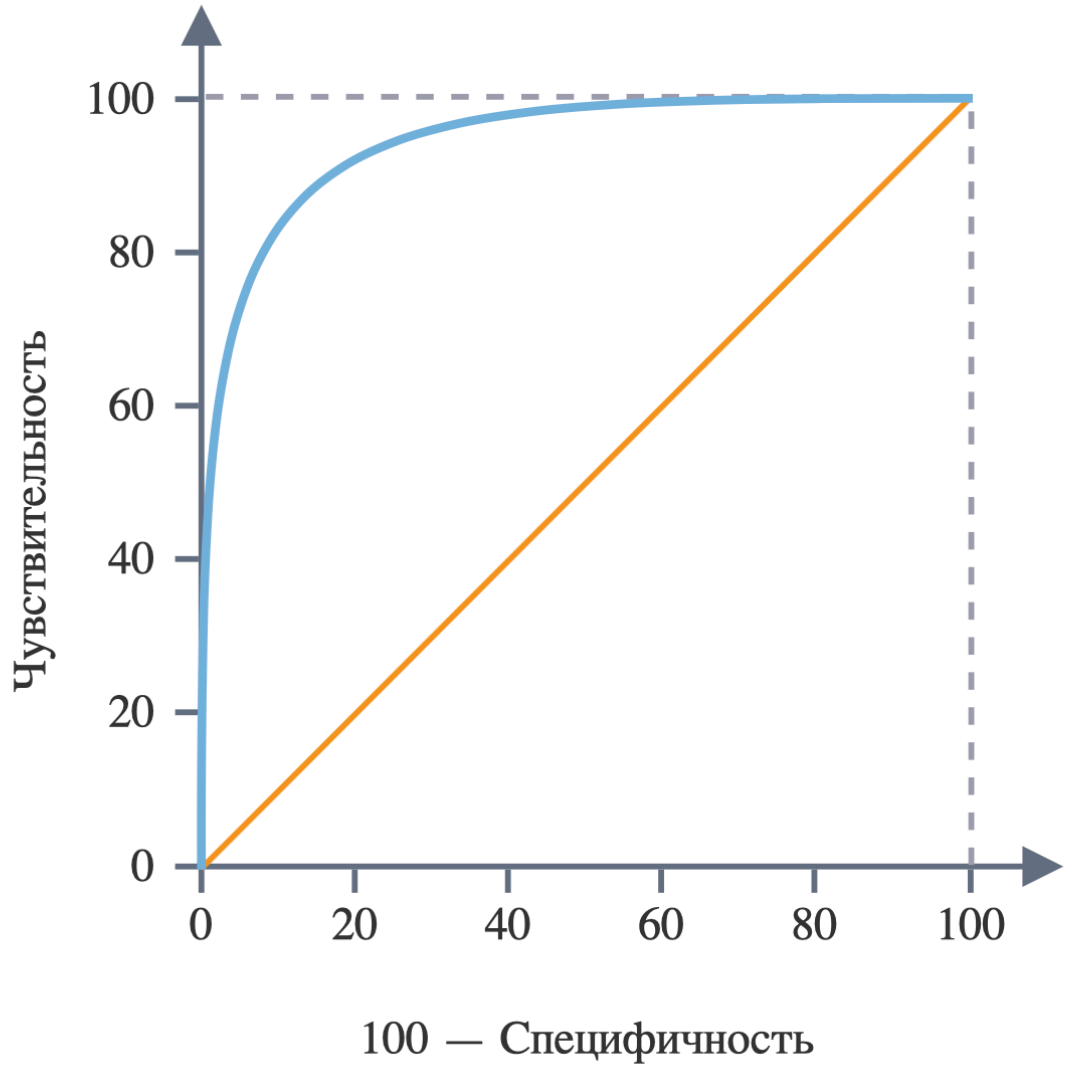
$$recall = \frac{TP}{TP + FN} = \frac{90}{90 + 10} = 0.9$$



Метрики бинарной классификации

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

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



Мультикласс

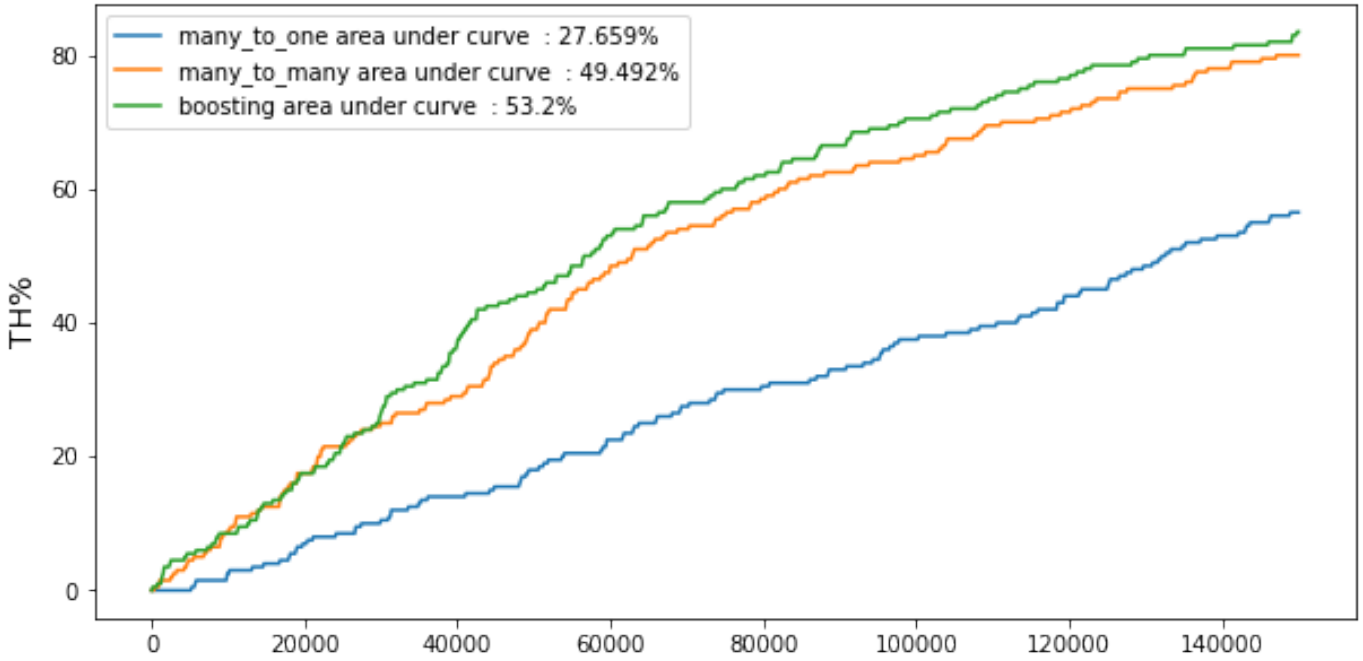
Actual Values	Predicted Values: Setosa	Predicted Values: Versicolor	Predicted Values: Virginica
Setosa	16 (cell 1)	0 (cell 2)	0 (cell 3)
Versicolor	0 (cell 4)	17 (cell 5)	1 (cell 6)
Virginica	0 (cell 7)	0 (cell 8)	11 (cell 9)

Метрики регрессии

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

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

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



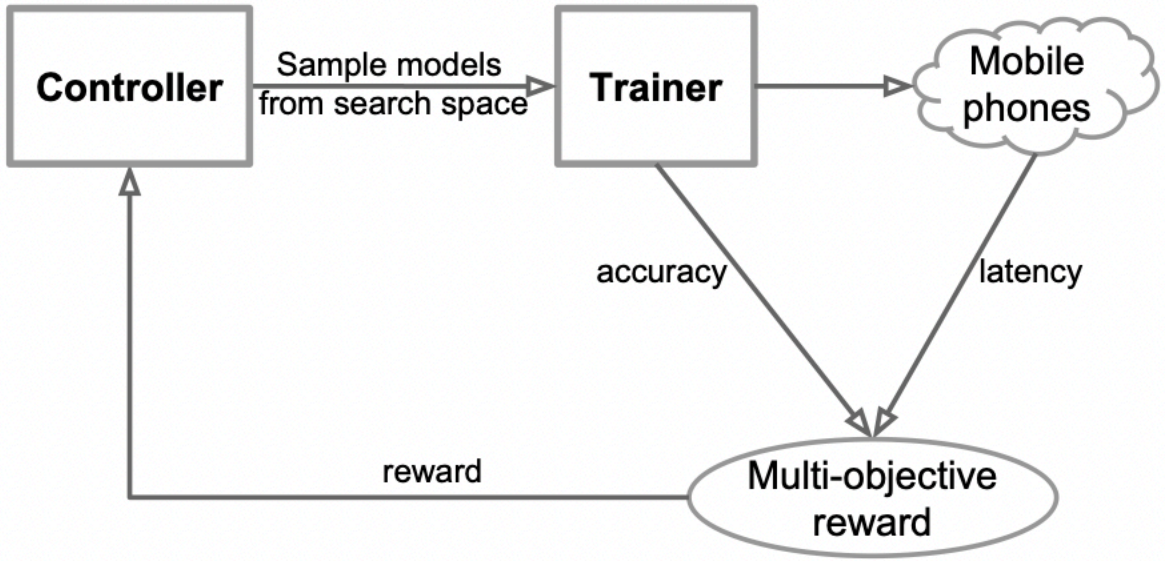


Figure 1: **An Overview of Platform-Aware Neural Architecture Search for Mobile.**

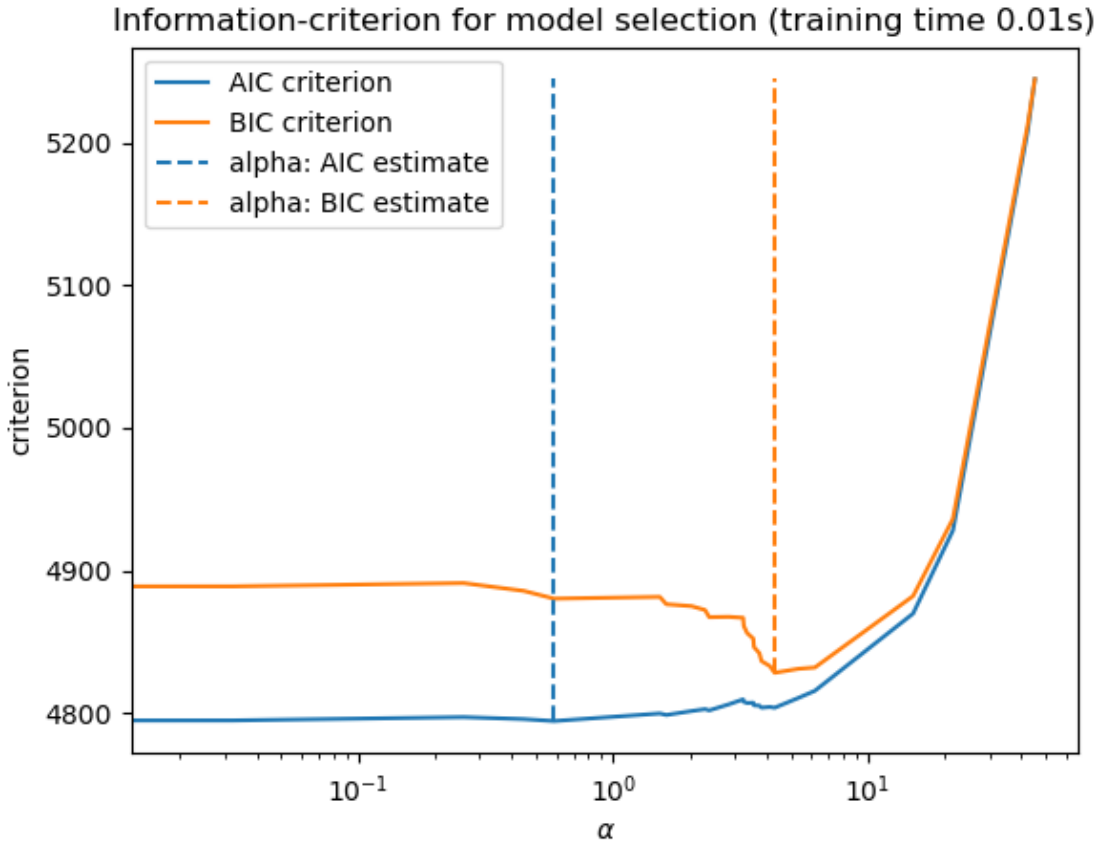
$$\underset{m}{\text{maximize}} \quad ACC(m) \times \left[\frac{LAT(m)}{T}\right]^{w}$$

https://arxiv.org/pdf/1807.11626.pdf

AIC/BIC

$$AIC = 2k - 2ln(L)$$

$$BIC = -2\ln(L) + k\ln(n)$$



https://machinelearningmastery.com/probabilistic-model-selection-measures/

https://scikit-learn.org/stable/auto_examples/linear_model/plot_lasso_model_selection.html

Критерий Стьюдента

- Параметрические
- Непараметрические

выборки: $X_1^{n_1} = (X_{11}, \ldots, X_{1n_1}), X_1 \sim N\left(\mu_1, \sigma_1^2\right)$
$X_2^{n_2} = (X_{21}, \ldots, X_{2n_2}), X_2 \sim N\left(\mu_2, \sigma_2^2\right)$

σ_1, σ_2 неизвестны

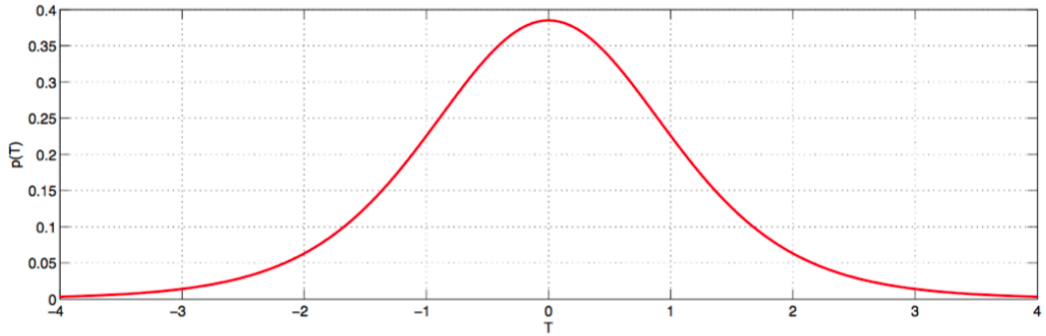
нулевая гипотеза: $H_0 : \mu_1 = \mu_2$

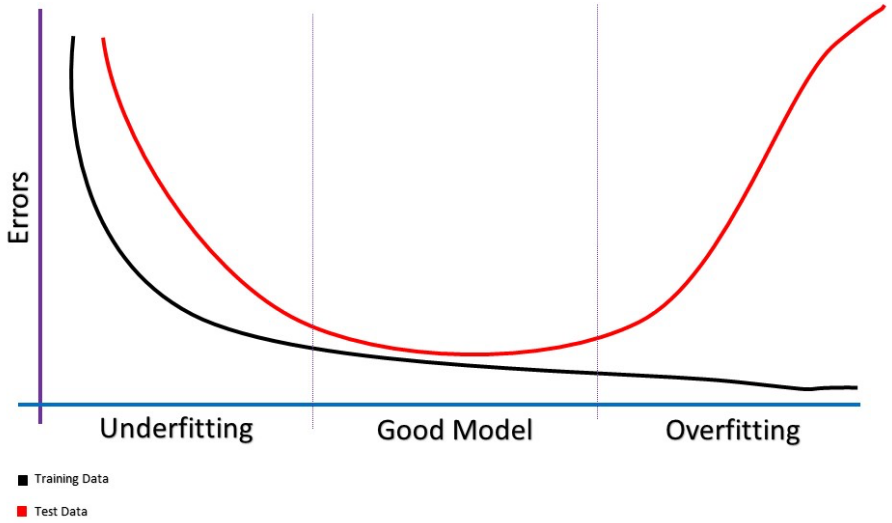
альтернатива: $H_1 : \mu_1 <\neq> \mu_2$

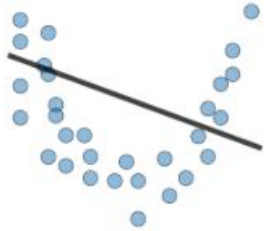
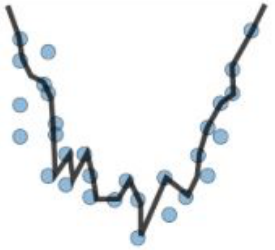
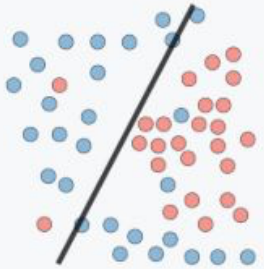
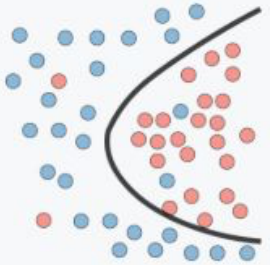
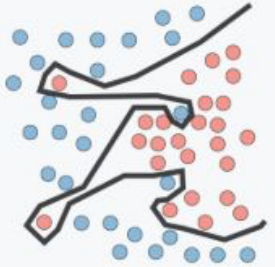
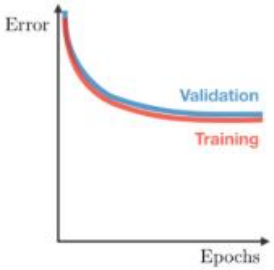
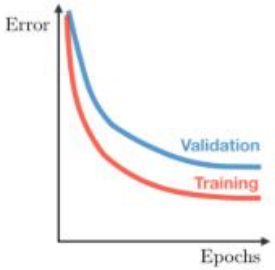
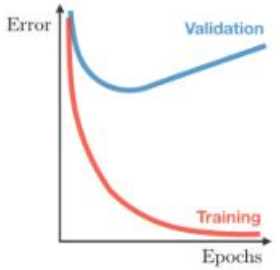
статистика: $T\left(X_1^{n_1}, X_2^{n_2}\right) = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}}$$

нулевое распределение: $\approx St(\nu)$





	Underfitting	Just right	Overfitting
Symptoms	• High training error • Training error close to test error • High bias	• Training error slightly lower than test error	• Very low training error • Training error much lower than test error • High variance
Regression illustration			
Classification illustration			
Deep learning illustration			
Possible remedies	• Complexify model • Add more features • Train longer		• Perform regularization • Get more data