

Калибровка моделей и uncertainty estimation

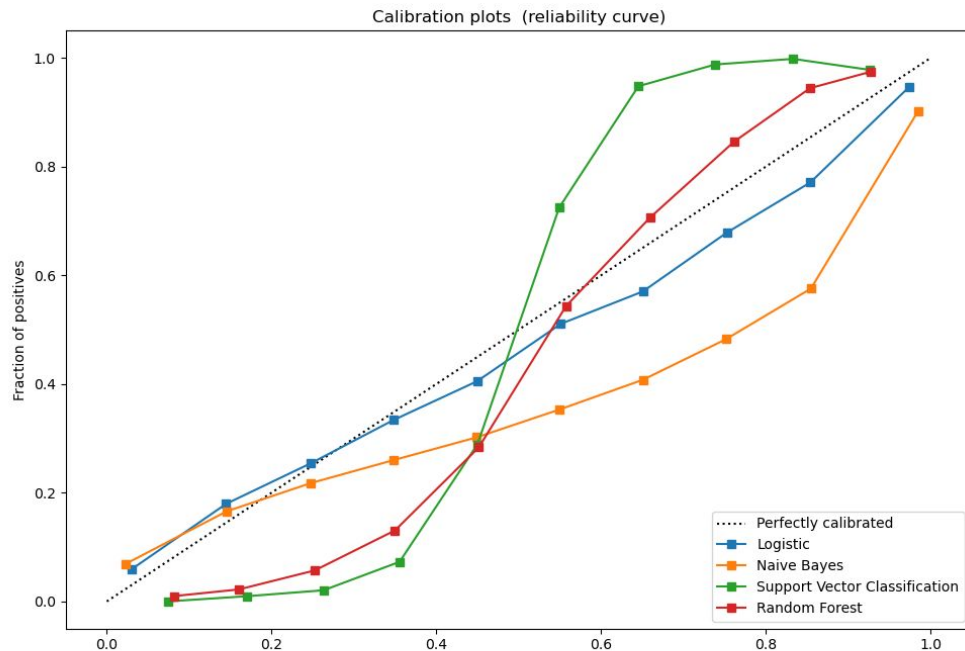
Терехова Юлия

Зачем калибровка?

модель 1	90%	99%
модель 2	90%	89%

- отражает, насколько классификация вероятна
- интерпретируемость модели

Что называется хорошо откалиброванной моделью?



$$P(y = +1 \mid b(x) = p) = p$$

Как оценить неоткалиброванность модели?

Expected Calibration Error (ECE)

$$\mathbb{E}_{\hat{P}} [|\mathbb{P}(\hat{Y} = Y | \hat{P} = p) - p|]$$

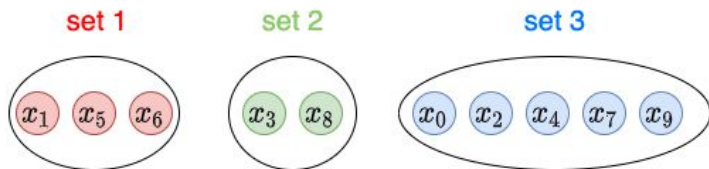
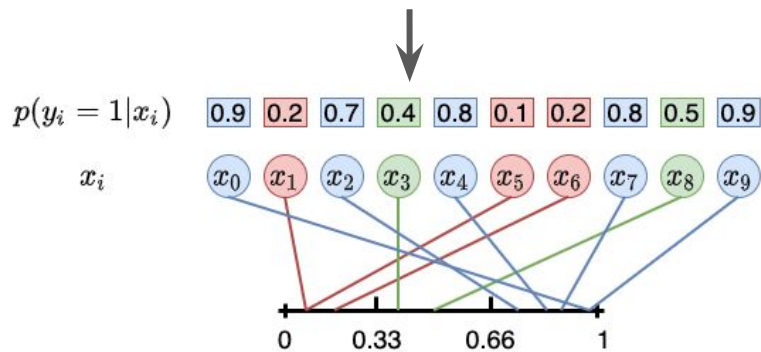
$$\text{ECE} = \sum_{m=1}^M \frac{|B_m|}{n} |\text{acc}(B_m) - \text{conf}(B_m)|$$

Maximum Calibration Error (MCE)

$$\max_{p \in [0,1]} |\mathbb{P}(\hat{Y} = Y | \hat{P} = p) - p|$$

$$\text{MCE} = \max_{m \in \{1, \dots, M\}} |\text{acc}(B_m) - \text{conf}(B_m)|$$

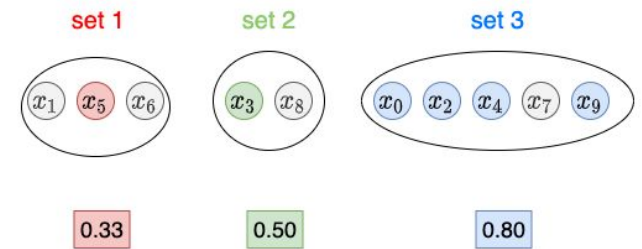
x_i	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
$p(y_i = 0 x_i)$	0.1	0.8	0.3	0.6	0.2	0.9	0.8	0.2	0.5	0.1
$p(y_i = 1 x_i)$	0.9	0.2	0.7	0.4	0.8	0.1	0.2	0.8	0.5	0.9



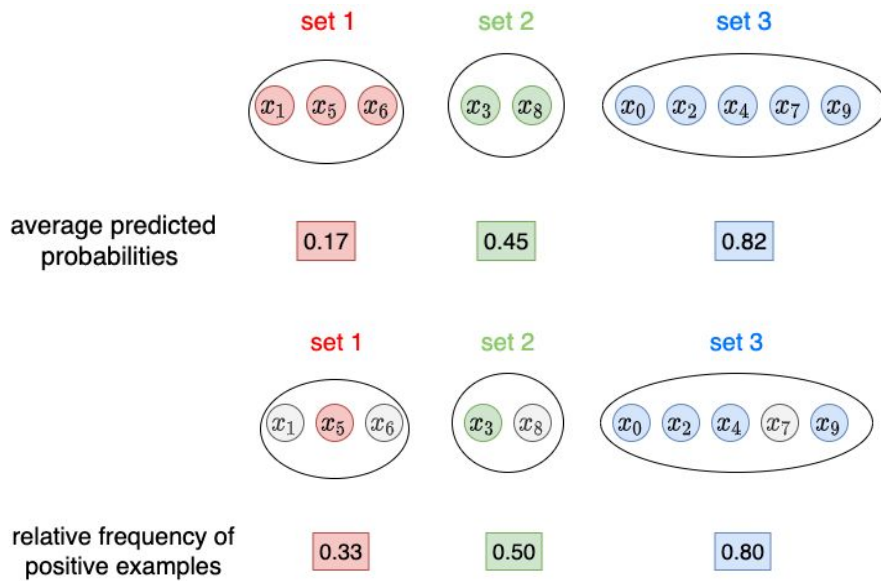
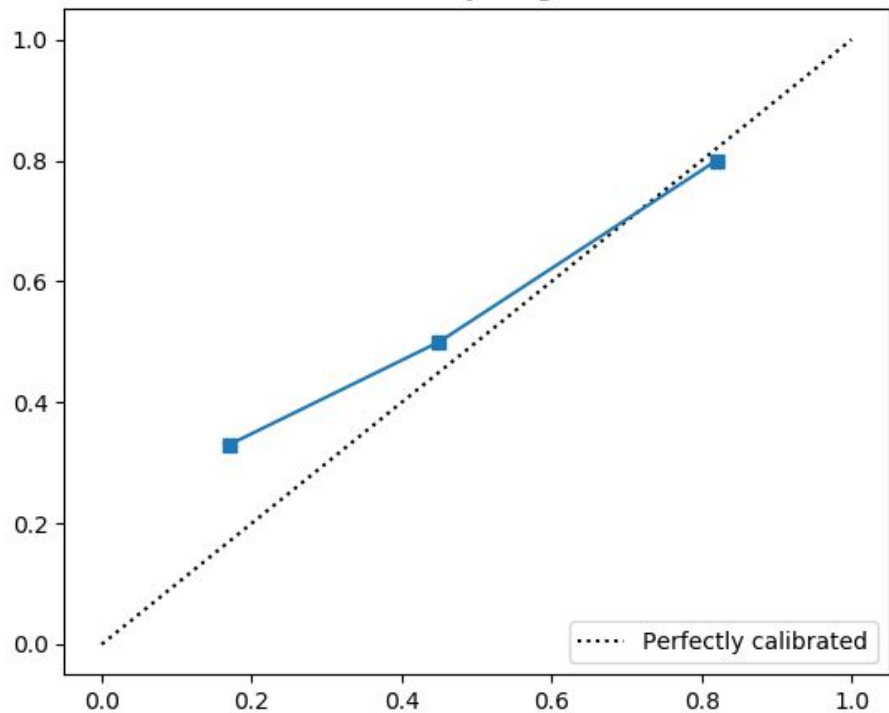
average predicted
probabilities



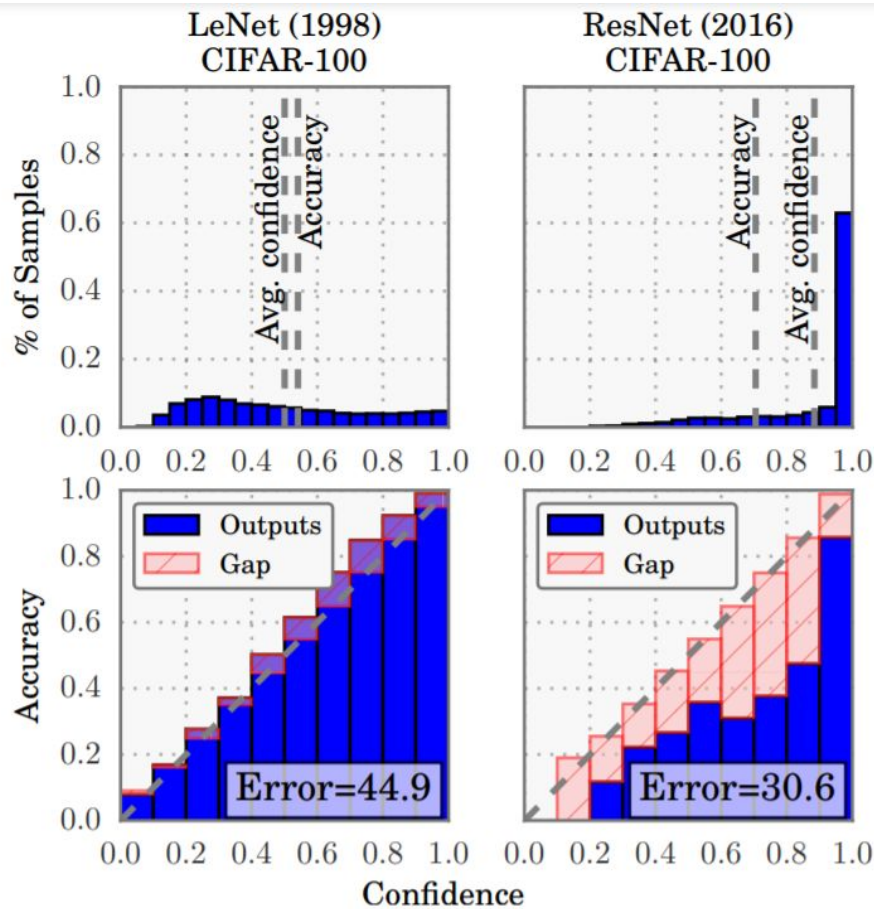
relative frequency of
positive examples



Reliability Diagram



Fun fact (net)



Platt scaling

$f(x)$ - output, чтобы получить
откалиброванные
вероятности - передать его в
сигмоиду

$$P(y = 1|f) = \frac{1}{1 + \exp(Af + B)}$$

$$\underset{A,B}{\operatorname{argmin}} \left\{ - \sum_i y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right\},$$

where

$$p_i = \frac{1}{1 + \exp(Af_i + B)}$$

Temperature scaling

От T не зависит
результат классификации

$$\hat{q}_i = \max_k \sigma_{\text{SM}}(\mathbf{z}_i/T)^{(k)}$$

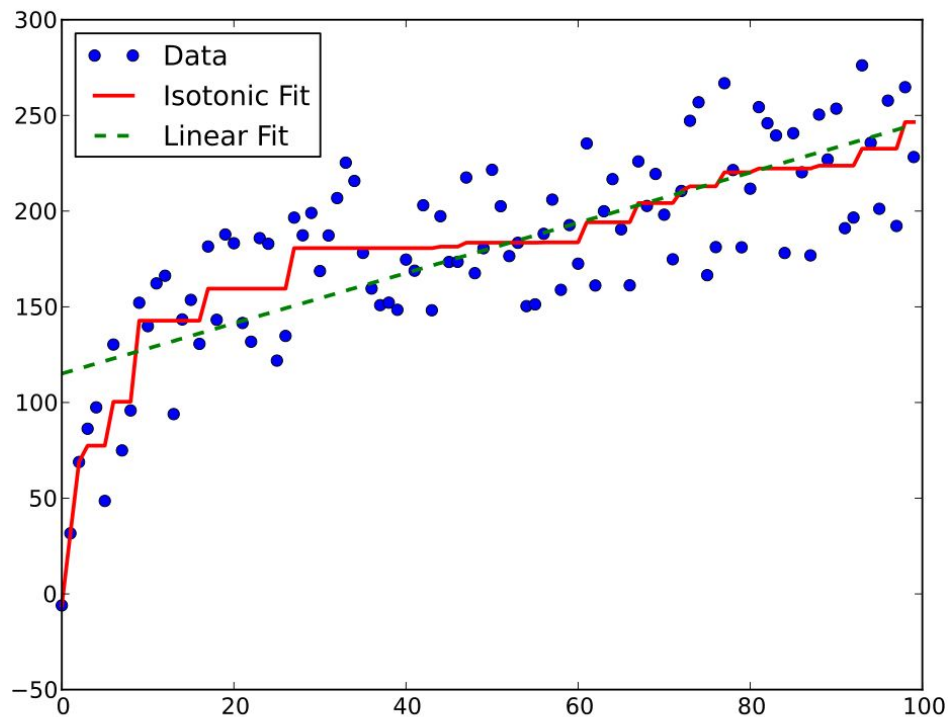
$$\sigma_{\text{SM}}(z) = \frac{e^{z^k}}{\sum_{k=1}^K e^{z^k}}$$

Isotonic regression

Непараметрический метод

$$y_i = m(f_i) + \epsilon_i$$

$$\hat{m} = \operatorname{argmin}_z \sum (y_i - z(f_i))^2$$



ИСТОЧНИКИ

[Predicting Good Probabilities With Supervised Learning](#)

[How and When to Use a Calibrated Classification Model with scikit-learn](#)

[On Calibration of Modern Neural Networks](#)