Калибровка моделей и uncertainty estimation

Смирнов Тимофей

Как выдавать предикт

$$b(x) = (b_1(x), \ldots, b_l(x)) \in [0,1]^l_{|\cdot|_1=1},$$
$$a(x) = \arg\max_j (b_j).$$

Вероятность верности предикта

$$b_{a(x)}(x) \approx \mathbf{P}(y(x) = a(x)).$$

Как это посчитать

$$\frac{1}{|B|}\sum_{x \in B} b_k(x) \text{ vs } \frac{1}{|B|}\sum_{x \in B} I[y(x) = k].$$

Диаграмма калибровки

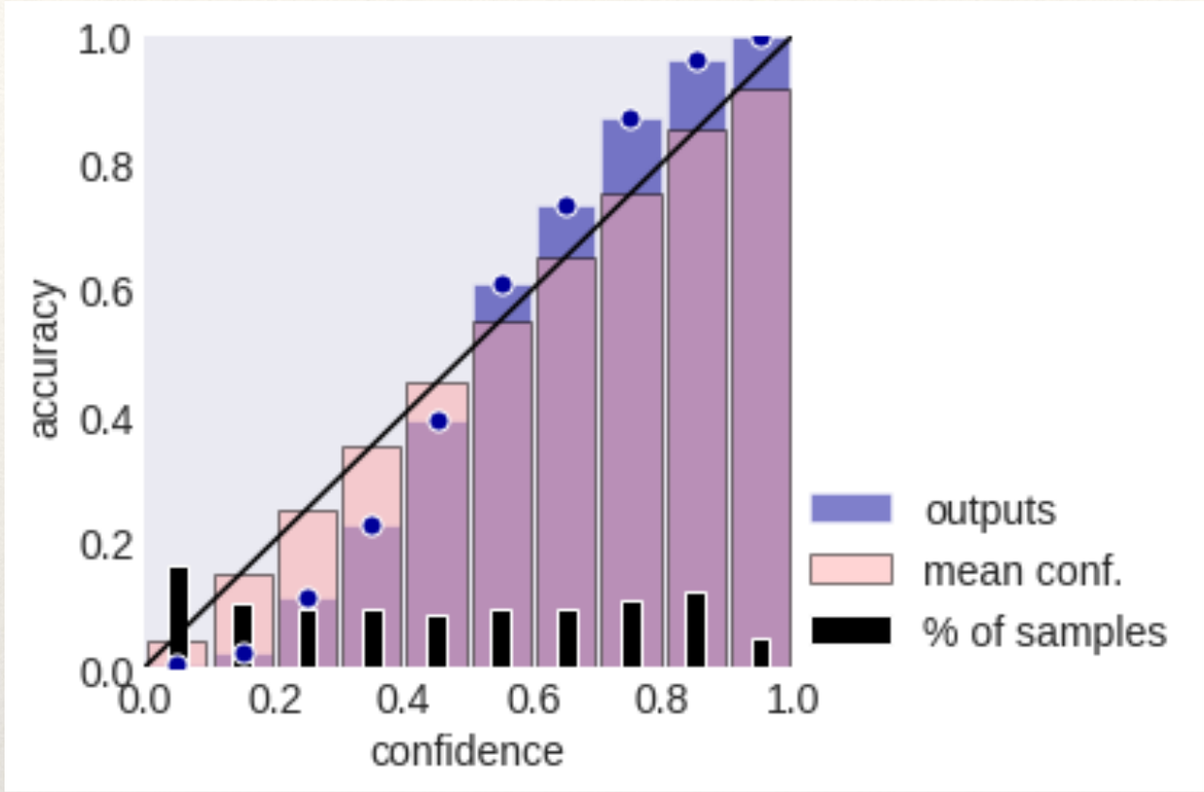
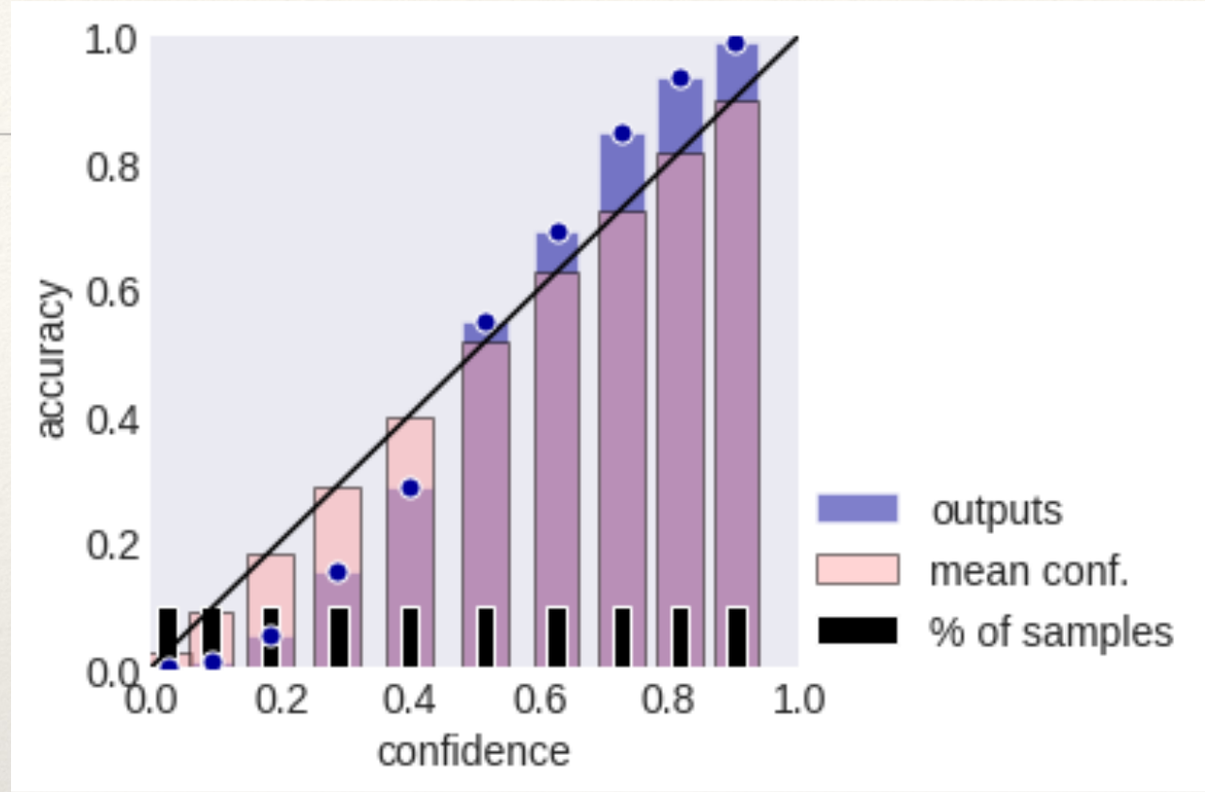


Диаграмма калибровки с равномощными бинами

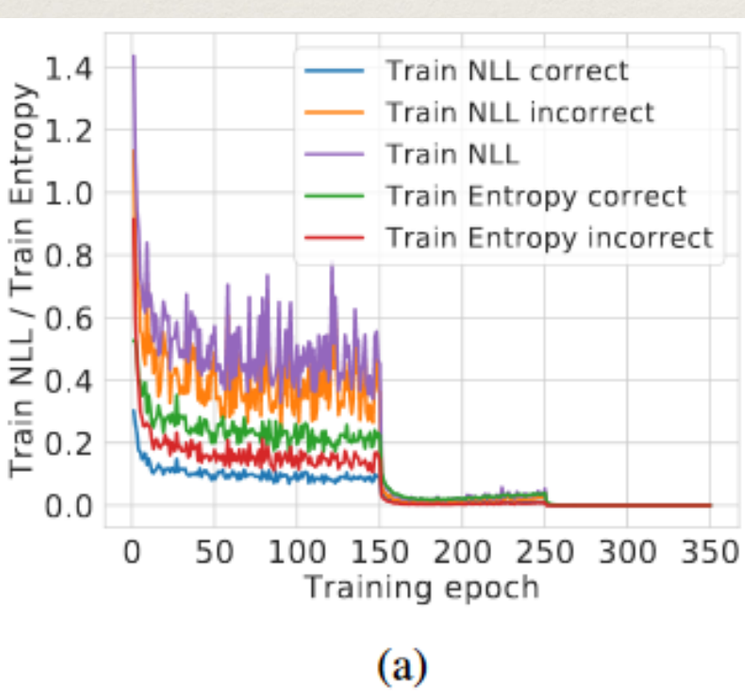


Expected Calibration Error

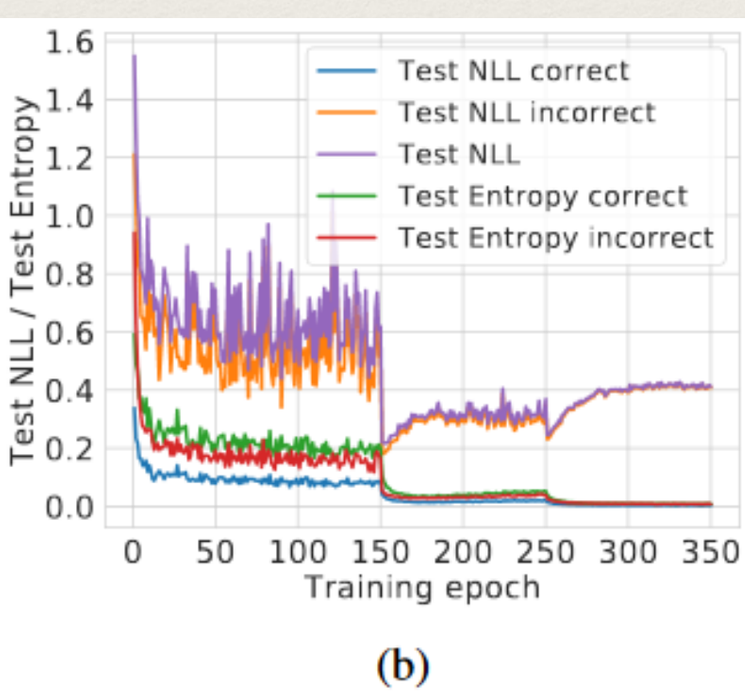
$$\frac{1}{|\{B\}|}\sum_{B\in\{B\}}\frac{|B|}{m}\left|\frac{1}{|B|}\sum_{x\in B}b_k(x)-\frac{1}{|B|}\sum_{x\in B}I[y(x)=k]\right|=$$

$$=\frac{1}{|\{B\}|\,m}\sum_{B\in\{B\}}\left|\sum_{x\in B}b_k(x)-\sum_{x\in B}I[y(x)=k]\right|$$

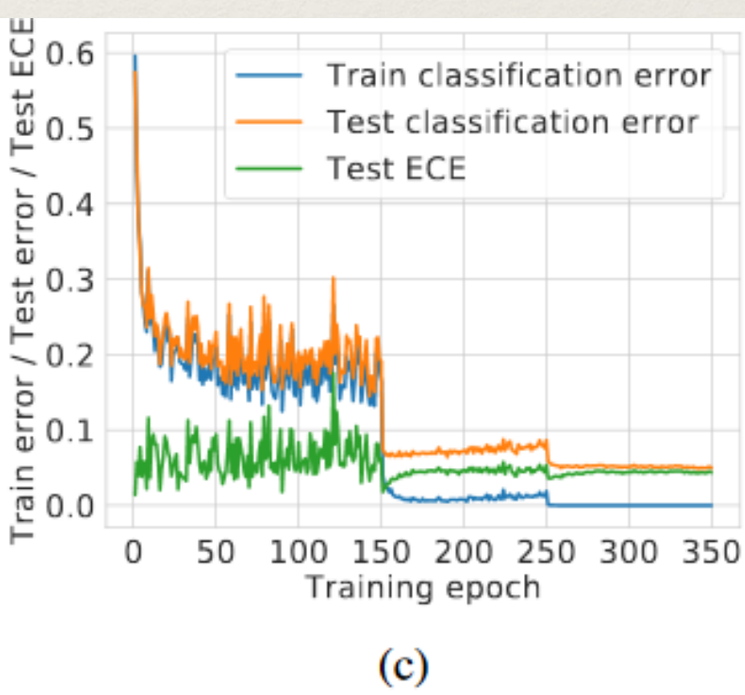
$$\text{NLL} = -\frac{1}{m}\sum_{j=1}^{m} y_j \log b_j \qquad \text{MSE} = -\frac{1}{m}\sum_{j=1}^{m} (y_j - b_j)^2$$



(a)

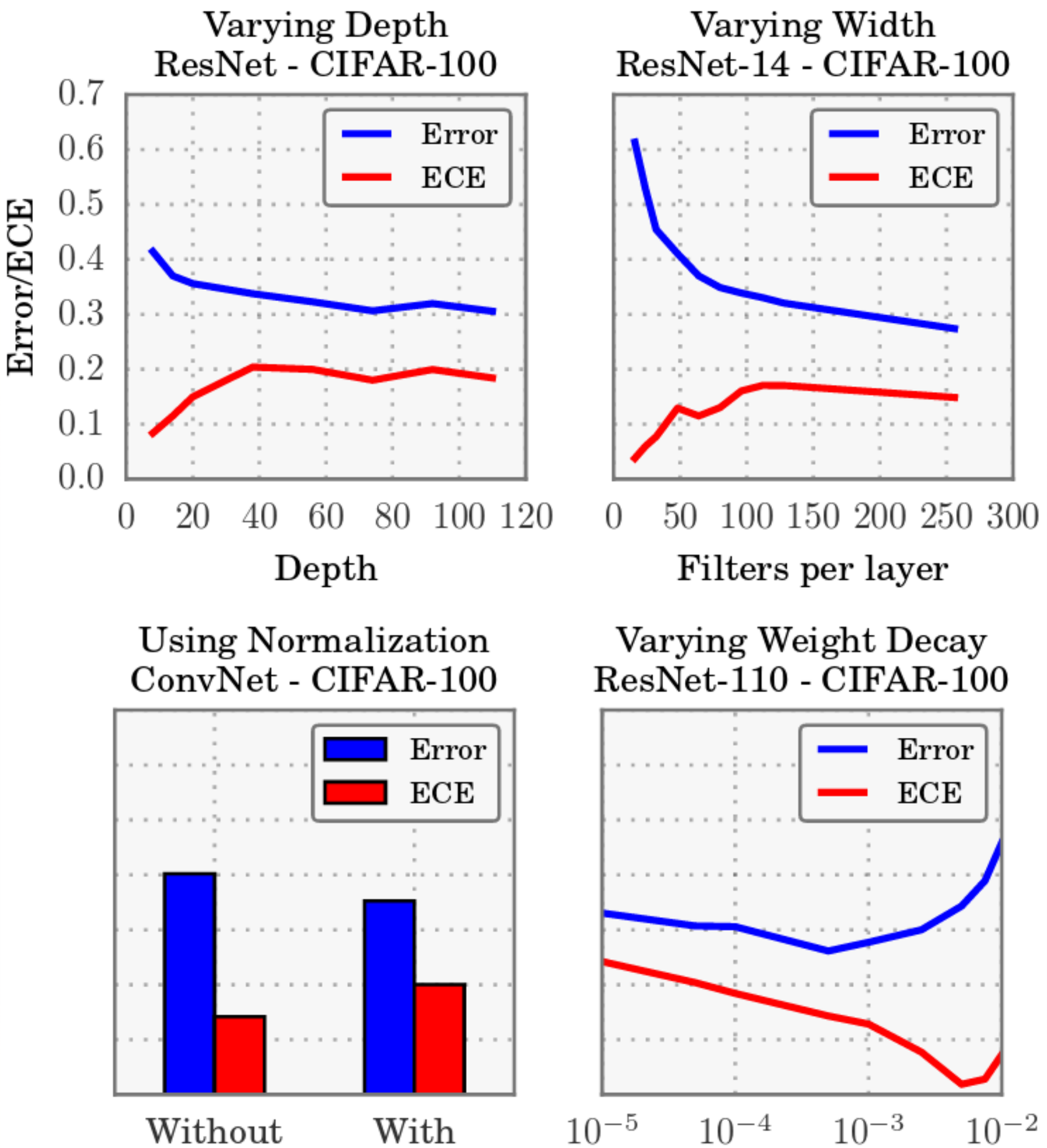


(b)

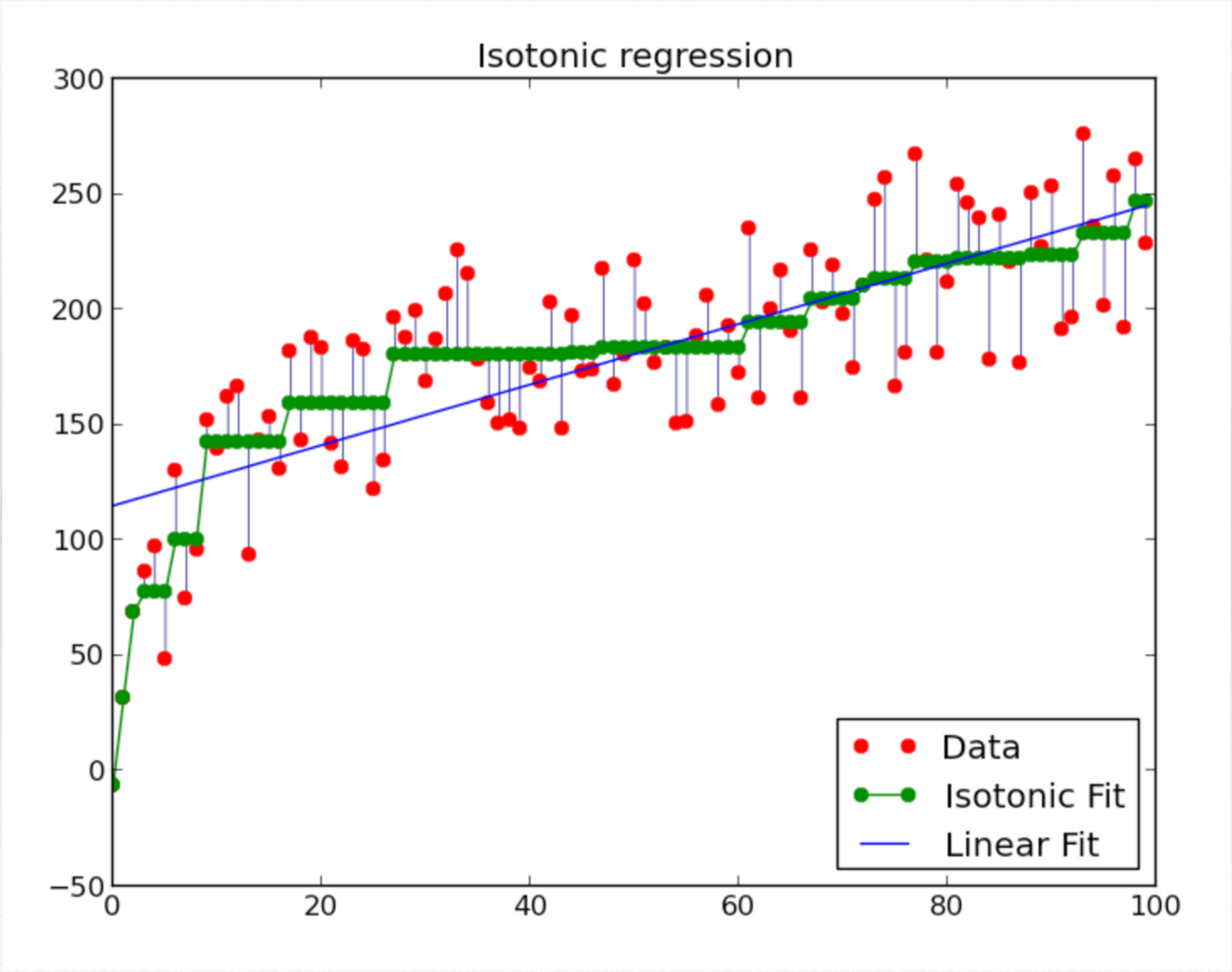


(c)

Figure 2: How metrics related to model calibration change whilst training a ResNet-50 network on CIFAR-10.

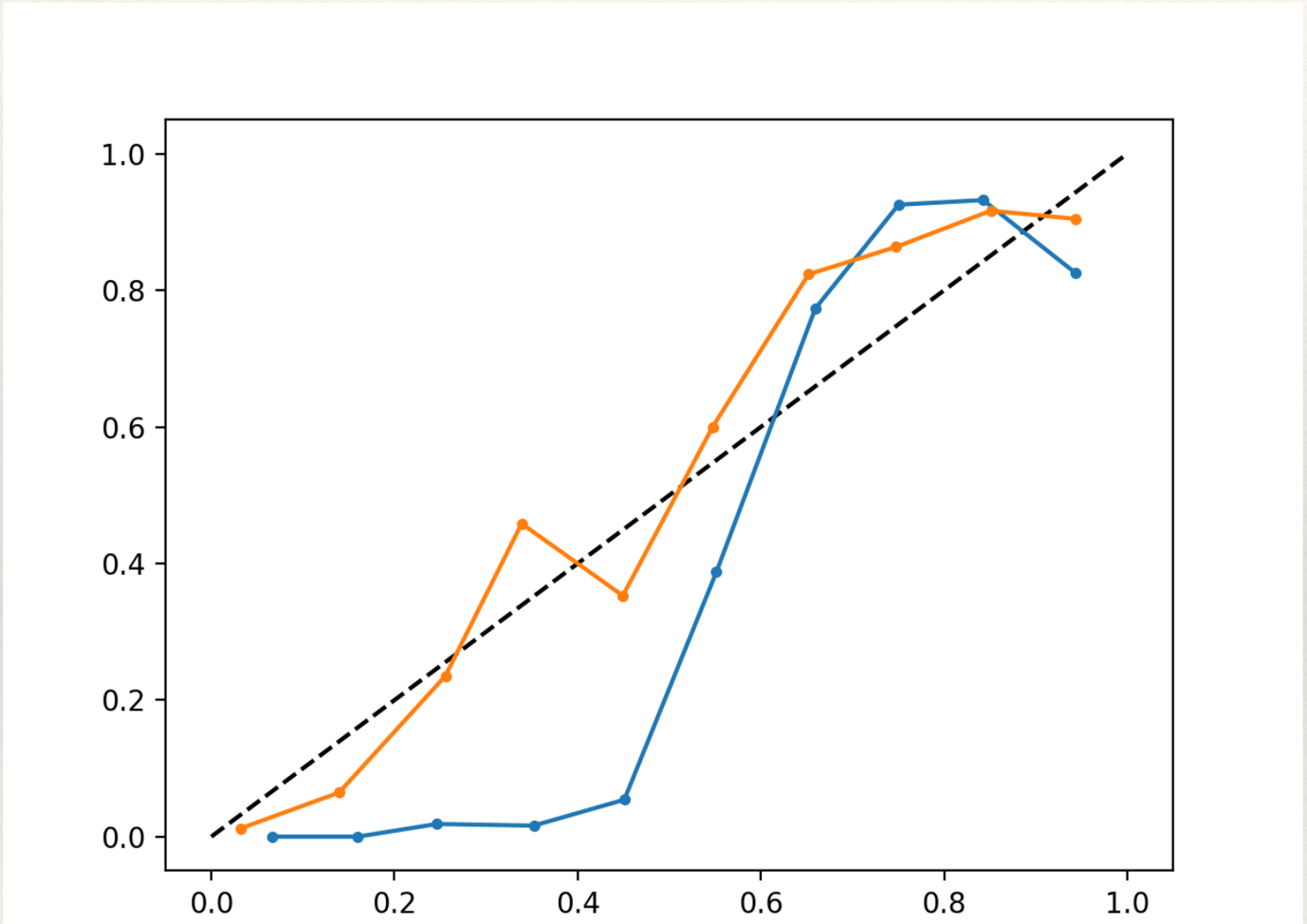

Varying Depth
ResNet - CIFAR-100
Error/ECE
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Error
ECE
0
20
40
60
80
100
120
Depth
Varying Width
ResNet-14 - CIFAR-100
Error
ECE
0
50
100
150
200
250
300
Filters per layer
Using Normalization
ConvNet - CIFAR-100
Error
ECE
Without
With
Batch Normalization
Varying Weight Decay
ResNet-110 - CIFAR-100
Error
ECE
10^{-5}
10^{-4}
10^{-3}
10^{-2}
Weight decay

Изотоническая регрессия



Калибровка Платта

$$b_{\text{new}}(x) = \text{sigmoid}(\alpha \cdot r(x) + \beta)$$

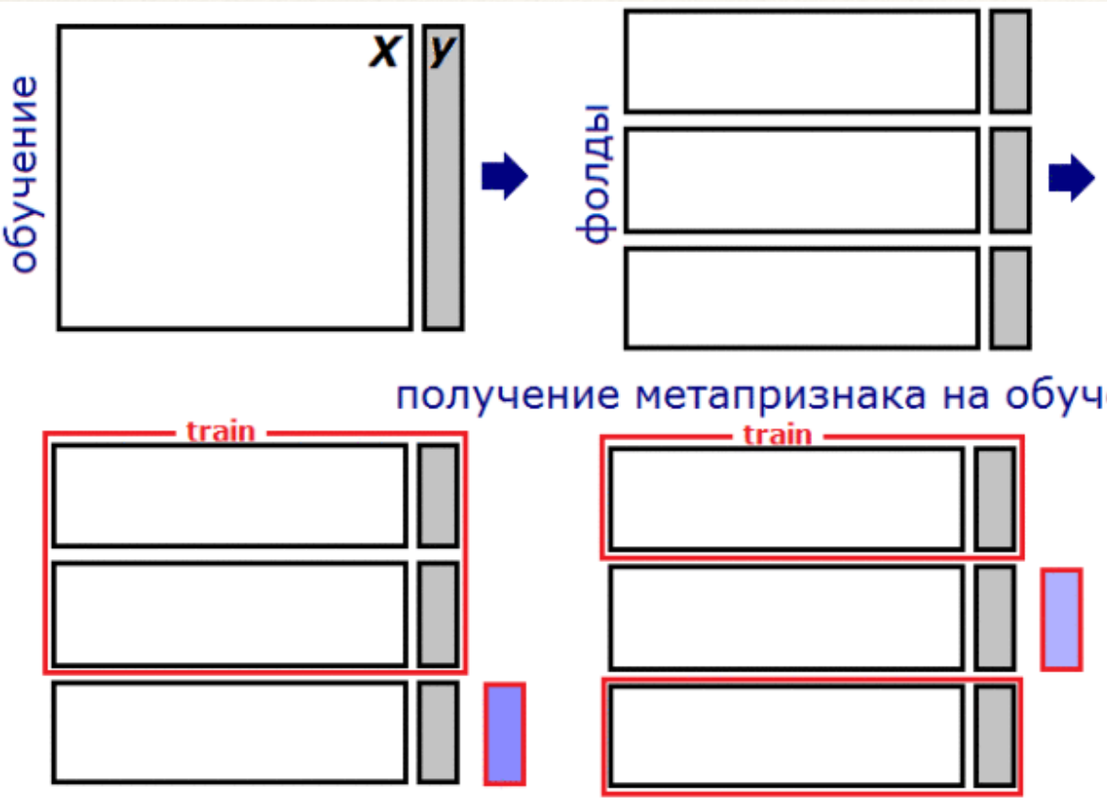


Температурное шкалирование $a(x) = \text{softmax}(b_1 / T, \ldots, b_l / T)$

Когда нужна калибровка ?

1. Для правильного понимания насколько можно доверять результатам модели(например человеческая обработка граничных случаев, краудсорсинг)
2. Для хорошего стакинга моделей(тк модели если модели по разному откалиброваны это может плохо сказаться на стакинге)
3. Когда на основе предиктов считается uplift или какая-то денежная выгода

Стакинг


обучение
X
y
фолды
получение метапризнака на обучении
train
train
train
получение метапризнака
на контроле
train
тест

Uplift-моделирование

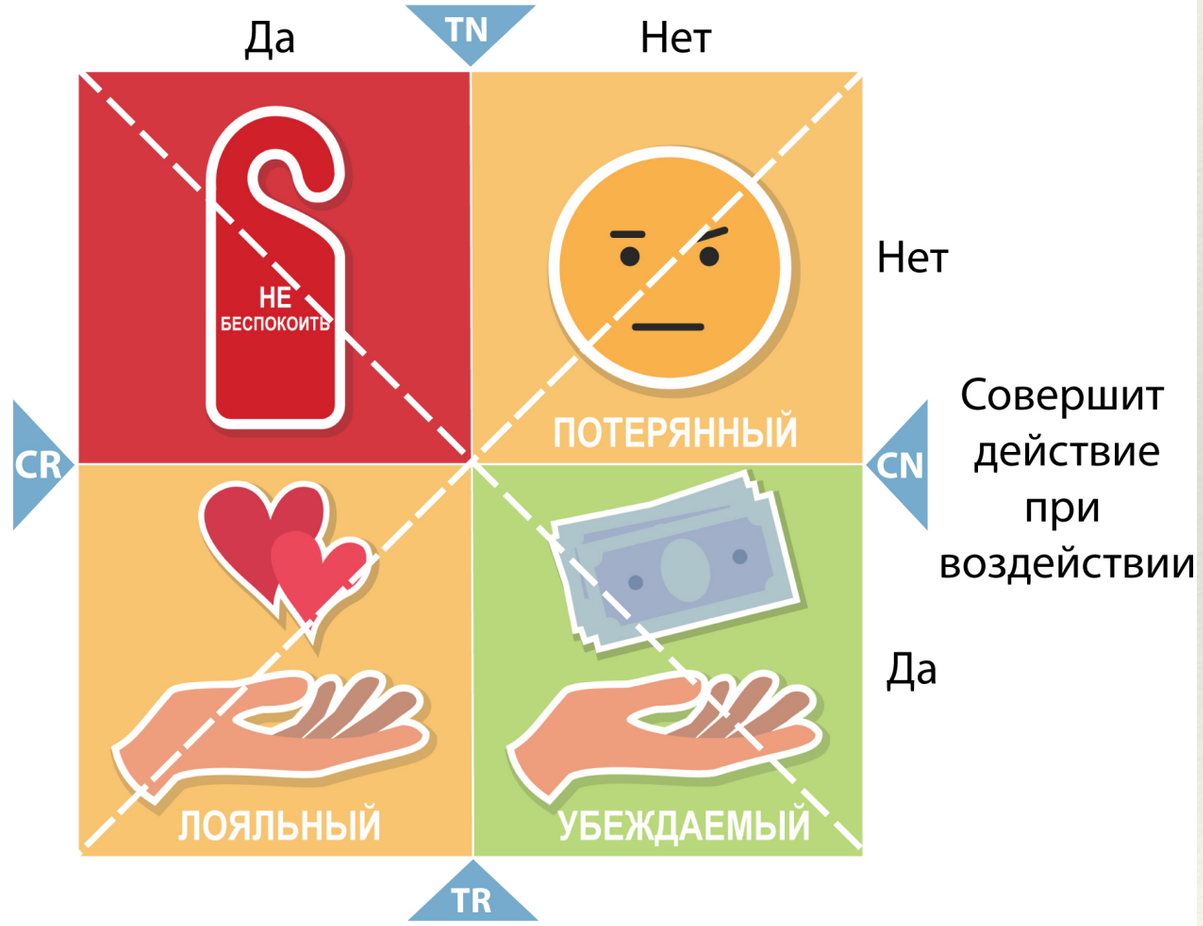
Что мы знаем:

Совершил действие

Было воздействие	Да	Нет
Нет	CR y=1 w=0	CN y=0 w=0
Да	TR y=1 w=1	TN y=0 w=1

Что мы хотим знать:

Совершит действие
без взаимодействия



Uplift-моделирование

Look-alike модель

P(целевого действия на основе схожести)

Response модель

P(целевого действия при коммуникации)

Uplift модель

P(целевого действия при коммуникации) — P(целевого действия без коммуникации)

Uplift * communication_cost **vs** marginality

Выводы

1. Калибровка нужна когда нам нужно интерпретировать результаты модели или использовать ее вероятности в бизнес метриках
2. Для нейросетей/SVM лучше всего использовать калибровку Платта/Температурное шкалирование