# Бутстрэппинг нейросетевых ансамблей

Шокоров Вячеслав Александрович

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Научный руководитель к.ф.-м.н. Д. П. Ветров

Москва,
2021 г.

# Ансамблирование

## Задача

Получение качественных, незашумленных данных для обучения нейронных сетей может достаточно дорого стоить.

## Проблема

1. При небольшом размере обучающей выборки многие модели в ансамбле получаются схожими, что увеличивает смещенность оценки итогового ансамбля.
2. Также различные подходы построения ансамбля позволяют получать различные оценки качества. Необходим единый универсальный метод построения ансамбля.

## Решение

1. Применение ансамблирования позволяет повысить точность и качество итоговой модели. Предлагается так же применить бутстрэппирование для обучения моделей внутри ансамбля.
2. Предлагается применение калиброванного логарифма правдоподобия.

# Литература


1. E. Lobacheva, N. Chirkova, M. Kodryan, D. P. Vetrov *On Power Laws in Deep Ensembles*, 2020.
2. A. Ashukha, A. Lyzhov, D. Molchanov, D. Vetrov. *Pitfalls of In-Domain Uncertainty Estimation and Ensembling in Deep Learning*, 2020.
3. J. Nixon, Tran *Why Aren't Bootstrapped Neural Networks Better*, 2021.

# Калиброванный логарифм правдоподобия

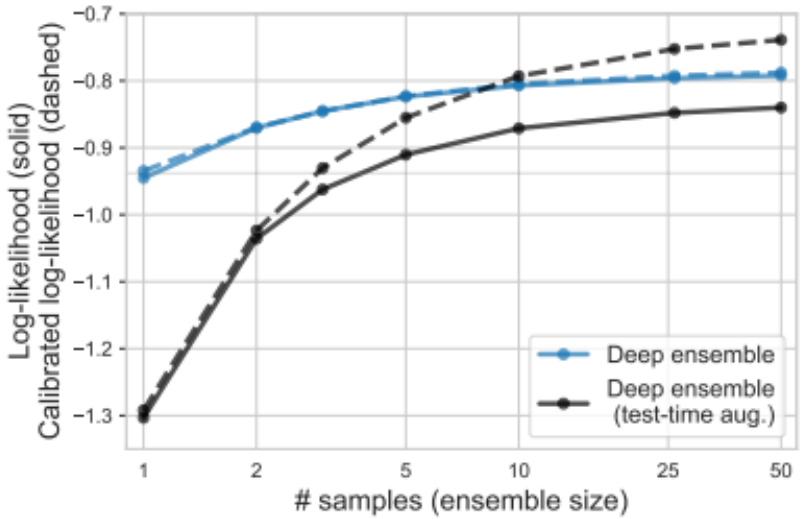



Figure 1: The average log-likelihood of two different ensembling techniques for ResNet50 on ImageNet dataset before (solid) and after (dashed) temperature scaling. Without the temperature scaling, test-time data augmentation decreases the log-likelihood of plain deep ensembles. However, when the temperature scaling is enabled, deep ensembles with test-time data augmentation outperform plain deep ensembles.

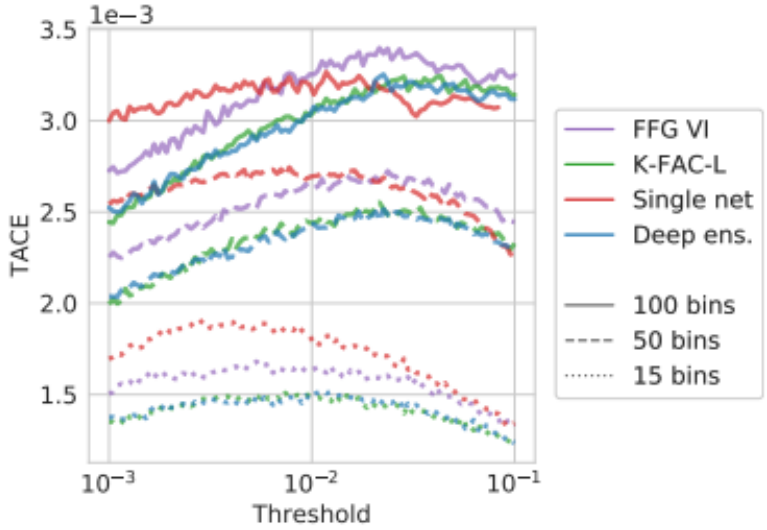
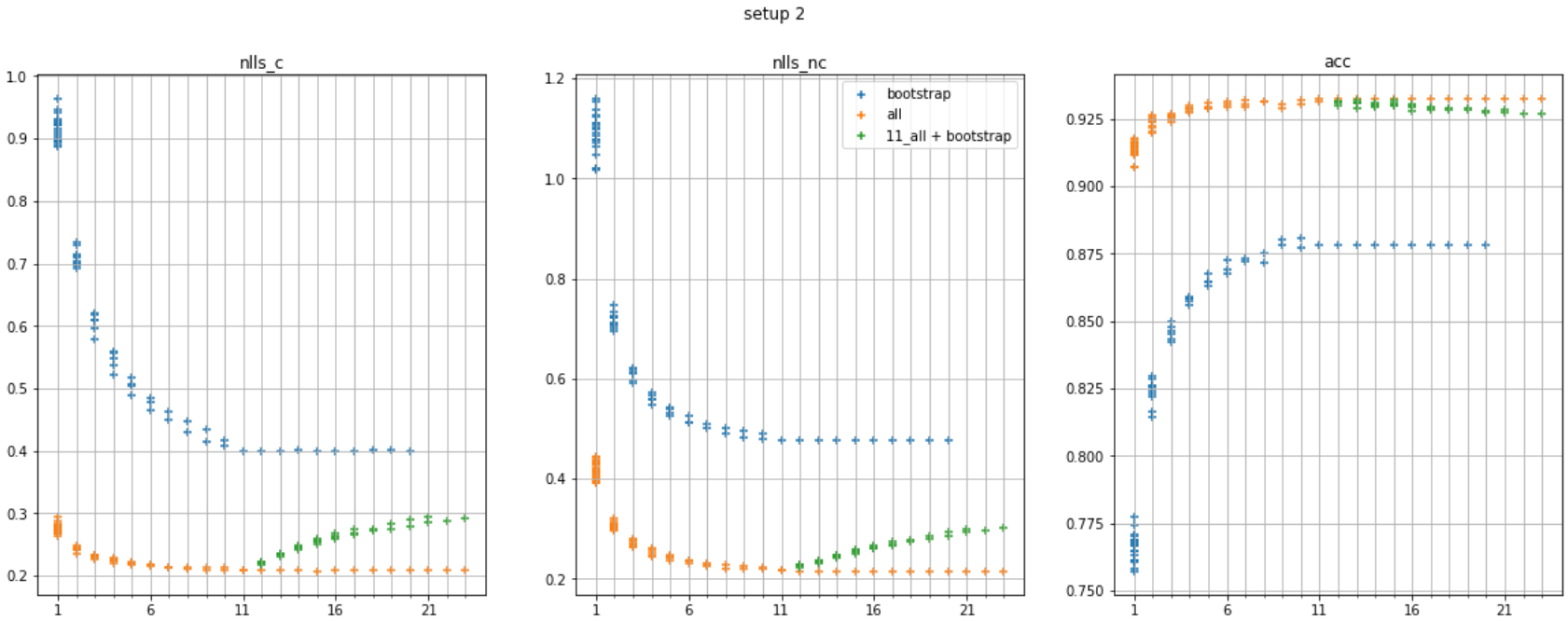



Figure 2: Thresholded adaptive calibration error (TACE) is highly sensitive to the threshold and the number of bins. It does not provide a consistent ranking of different ensembling techniques. Here TACE is reported for VGG16BN model on CIFAR-100 dataset and is evaluated at the optimal temperature.

# Калиброванный логарифм правдоподобия

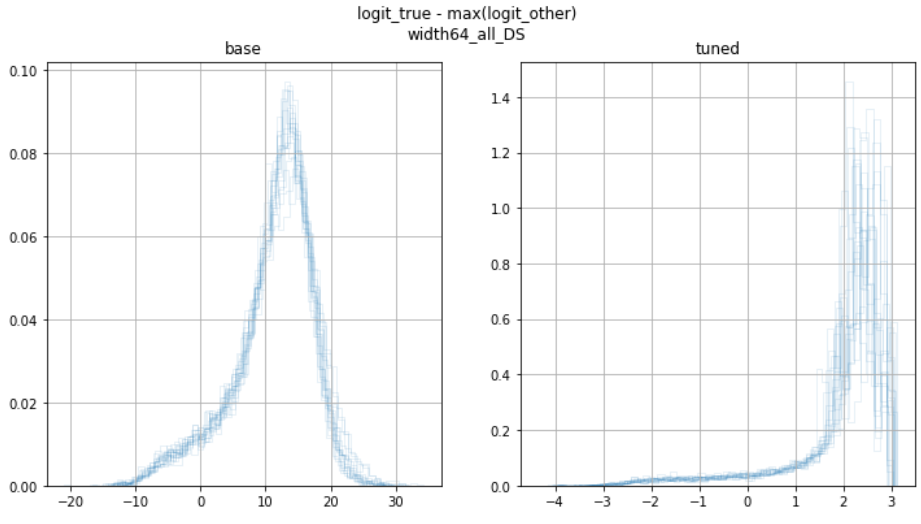
$$\mathrm{CNLL}_n = \mathbb{E} \min_{\tau>0} \left\{ - \sum_{\mathrm{obj} \in \mathcal{D}} \log \bar{p}^*_{\mathrm{obj},n}(\tau) \right\},$$

# Вычислительный эксперимент

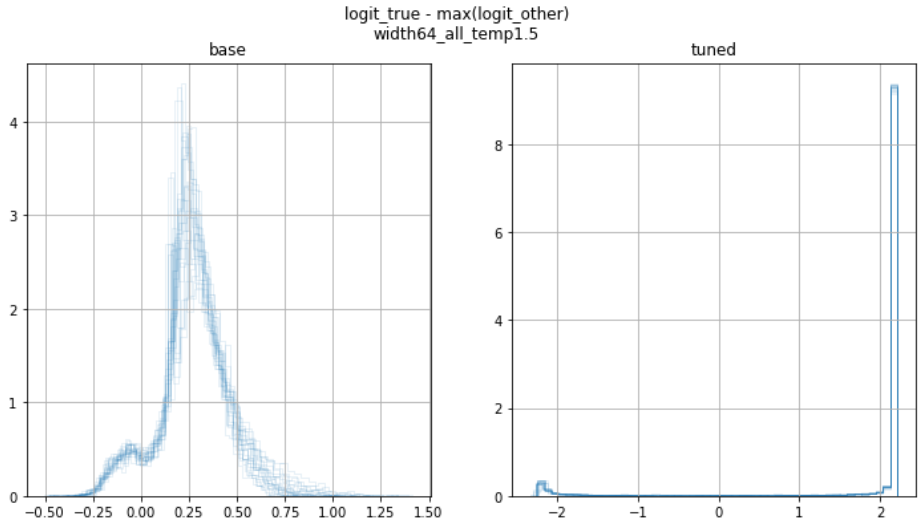


# Что такое зазор в логитах?

a)




b)

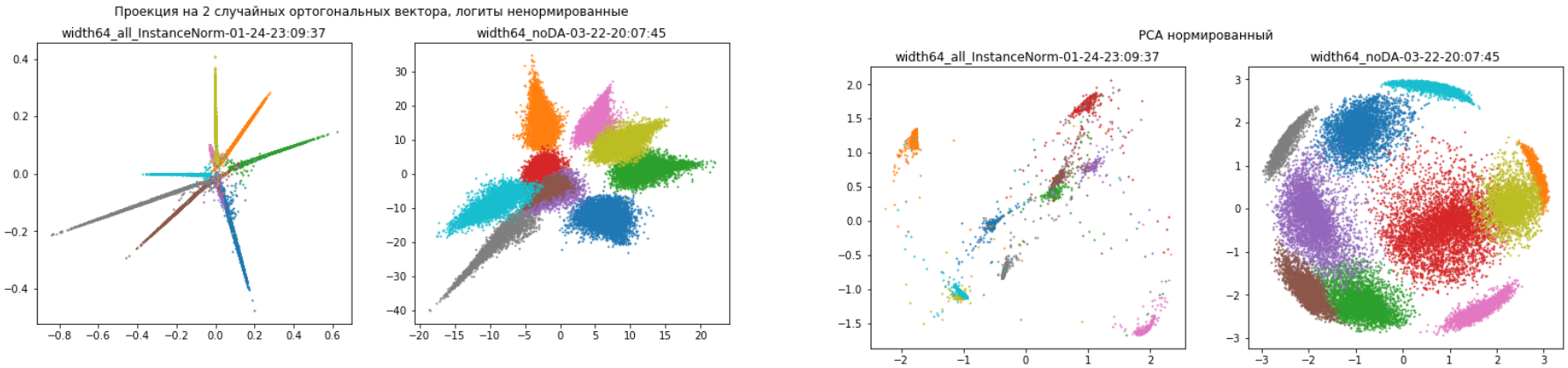



Пусть l - логиты модели, $y$ - метка правильного класса, тогда зазором называет значение $l_y - \max_{i \neq y} l_i$. На графиках изображены гистограммы зазоров для двух моделей. a) - для модели, которая обучалась стандартно, b) - для модели, которая обучалась при добавлении нормировки на логиты. Левый столбец соответствует изначальным логитам, правый - нормированным.

# Добавление нормировки в логитах

Почему это может быть интересно?

- Позволяет на этапе обучения заставить сеть увеличить зазор.




Если перейти в сферические координаты, у модели есть 2 возможности:

1. увеличить расстояние
2. уменьшить дисперсию по углу

Первый простой вариант, легко получить небольшое уменьшение лосса. Второй же более сложный вариант, требует от модели выучивание более сложных фичей.

Обучая модель на нормирующем функционале мы лишаем модель возможности 1, поэтому чтобы получить хорошее предсказание ей остается только 2.

# Обучение сети с гэпом

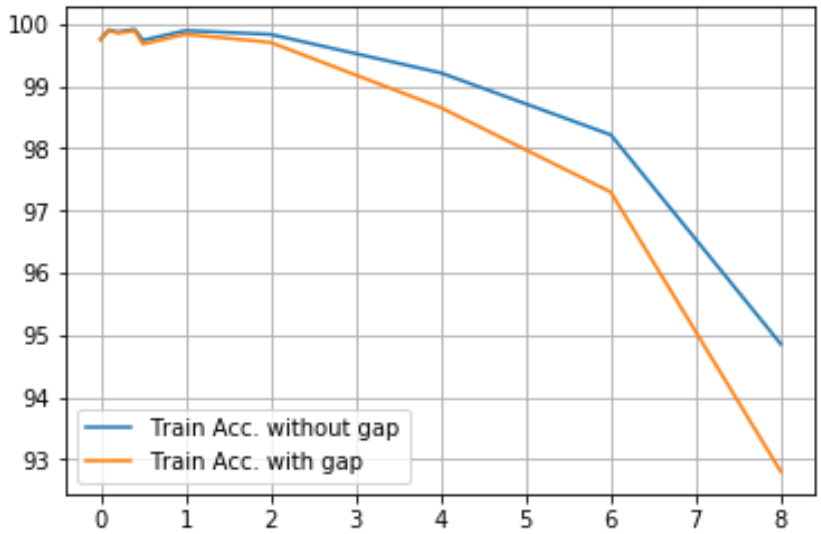
Пусть $\mathcal{L}(\mathsf{l}, y)$ - функция ошибки, $\mathsf{l}$ - значение логита на некотором объекте выборки, $y$ - метка правильного класса, тогда функция ошибки с гэпом будет:

$$\mathcal{L}_m = \mathcal{L}(\mathsf{l} - m\mathsf{e}_y, y),$$

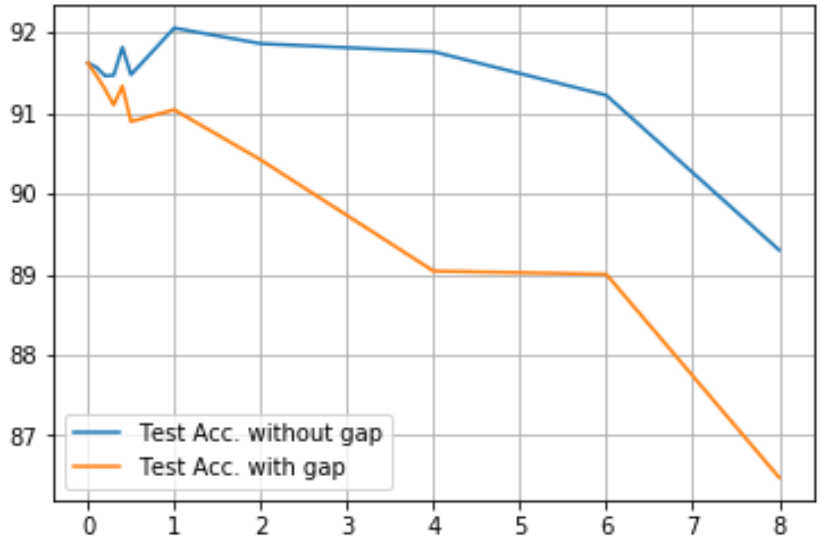
где $e_y$ - унитарный код с единицей на $y$-ой координате, $m$ - величина гэпа.

# Обучение сети с гэпом

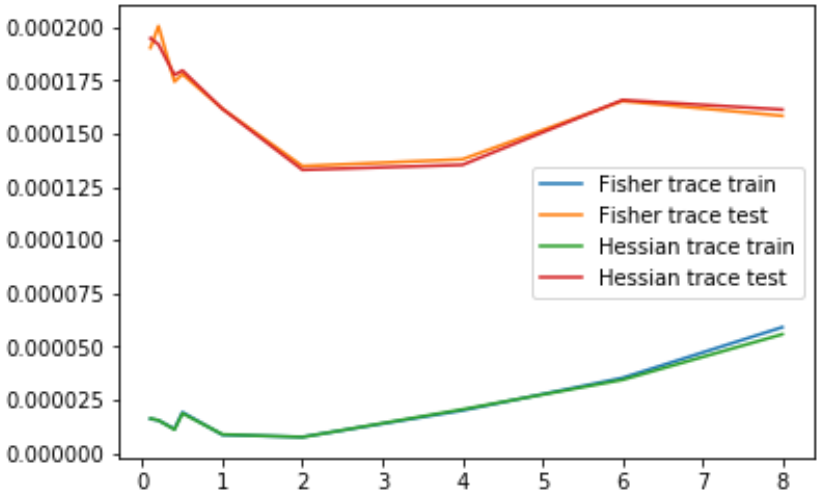
a)




b)




c)




Обучение сетей с гэпом позволяет повысить точность на обучении и тесте, а также позволяет увеличить ширину минимума, что говорит о повышении генерализации модели. a) - точность на обучающей выборке для различных размеров гэпа. b) - точность на тестовой выборке для различных размеров гэпа. c) - ширина минимум у полученных моделей. Ширина вычилслялась через след Гессиана и след матрицы Фишера.