

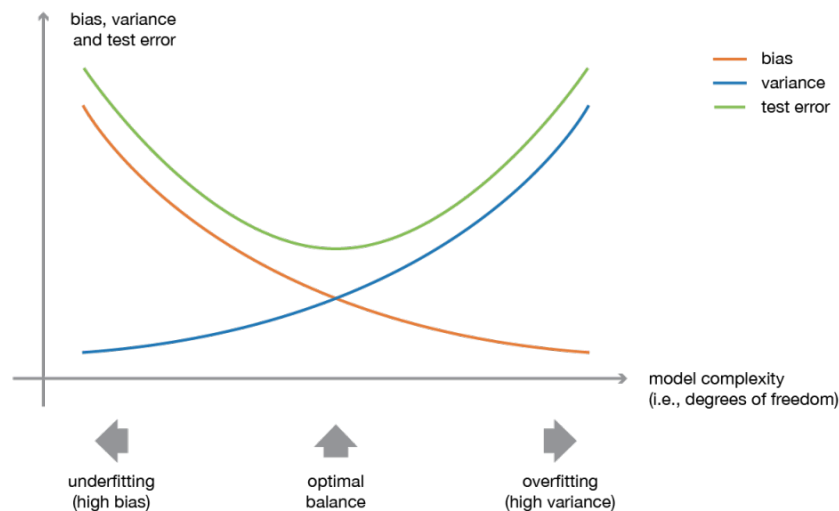
# Ансамблирование нейронных сетей

Студент БПМИ172 Ахметов Артемий

# Виды ошибок

В любую ошибку предсказания модели можно выделить три составляющие:

- Ошибка смещения
- Ошибка дисперсии
- Неуменьшаемая ошибка (шум)



# Ошибка смещения

Смещение - это предположения, сделанные моделью для облегчения изучения целевой функции. Соответственно:

- Низкое смещение - мало предположений о значении целевой функции (решающие деревья, KNN, SVM)
- Высокое смещение - делается много предположений о значении целевой функции (линейная регрессия, логистическая регрессия)

# Ошибка дисперсии

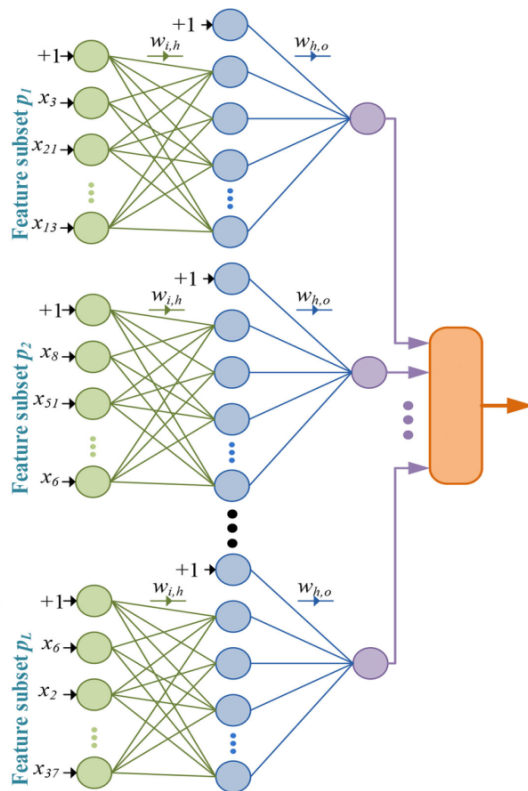
Предположение модели о том насколько будет меняться значение целевой функции.

- Низкая дисперсия - модель предполагает, что целевая функция мало меняется(линейная регрессия, логистическая регрессия)
- Высокая дисперсия - модель предполагает, что целевая функция сильно меняется(решающие деревья, KNN, SVM)

# Зачем ансамблируют нейронные сети

Ансамблирование нейронных сетей - обучение нескольких сетей на разных датасетах или разными методами и компоновка их результатов.

Идея ансамблей - разные модели не делают одинаковые ошибки. Таким образом уменьшаются ошибки смещения и дисперсии.



# Как ансамблируют нейронные сети

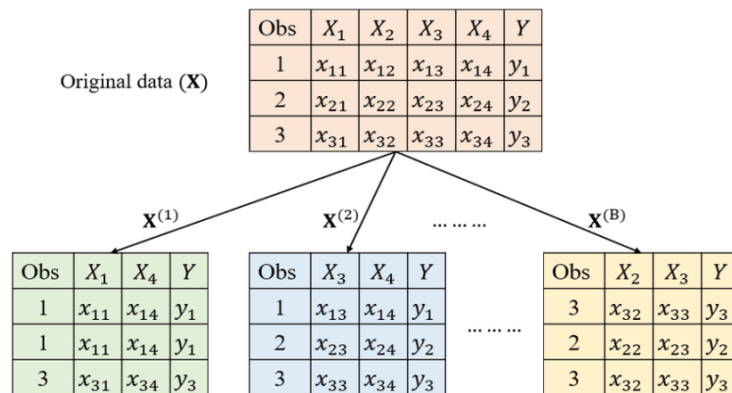
Ансамбль регулируется тремя параметрами:

- Модификации данных(resampling)
- Модификации моделей
- Компоновка результатов

# Ансамбли меняющие данные

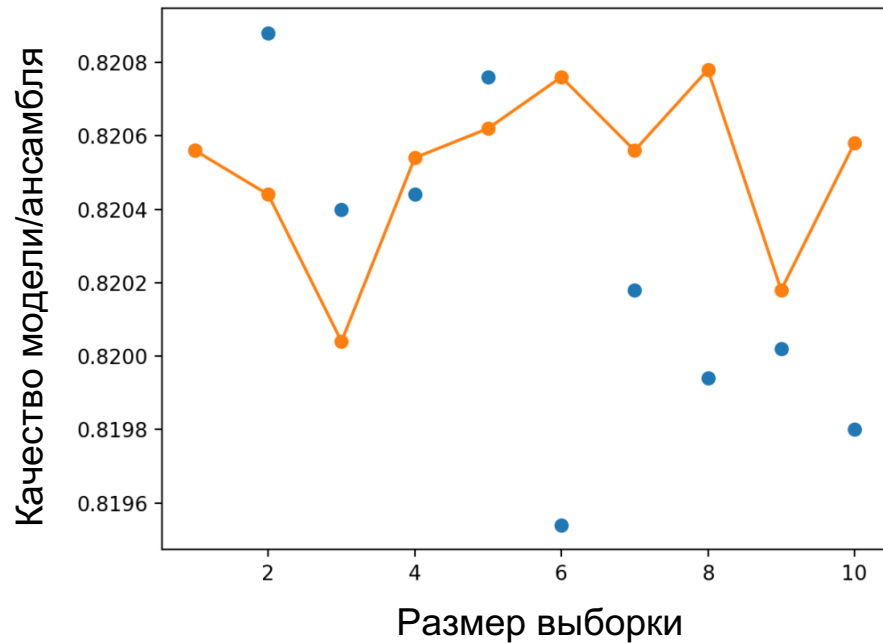
Один из подходов к созданию ансамблей - обучение одного алгоритма на разных датасетах. Вот некоторые из этих подходов:

- Random Splits Ensemble
- K-fold Cross-Validation Ensemble
- Bootstrap Aggregation(Bagging)



# Random Splits Ensemble

Простая генерация случайных частей из исходного датасета и обучение моделей на разных частях данных.

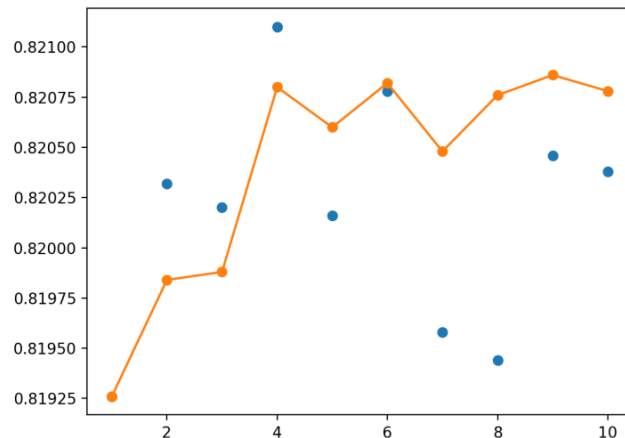




# K-fold Cross-Validation Ensemble

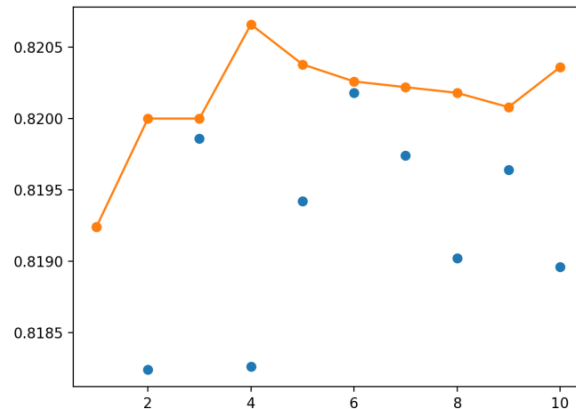
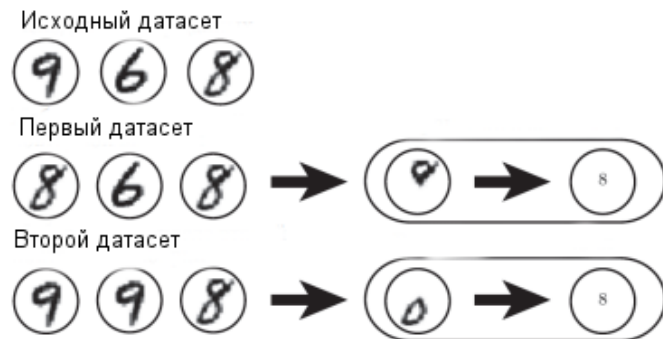
Весь датасет делится на  $k$  равных частей, на каждой из которых модель обучится  $k-1$  раз, и 1 раз часть будет тестовой. Метод не так смещен, так как каждая часть используется при обучении.

1. Перемешать данные
2. Разделить данные на  $k$  групп
3. Для каждой из групп
  - a. Модель обучается на остальных группах
  - b. Модель оценивается на тестовом датасете
  - c. Запоминается результат модели
4. Компонуется результат работы всех моделей



# Bagging(Bootstrap Aggregation)

Бутстреп - процедура генерации повторных случайных выборок из исходного набора данных. Бутстреп-выборки производятся равномерно и с возвращением, поэтому некоторые исходные примеры будут отсутствовать, а другие - дублироваться. Немного страдает от смещения, но в целом не хуже k-fold Cross Validation.



# Ансамбли меняющие модели

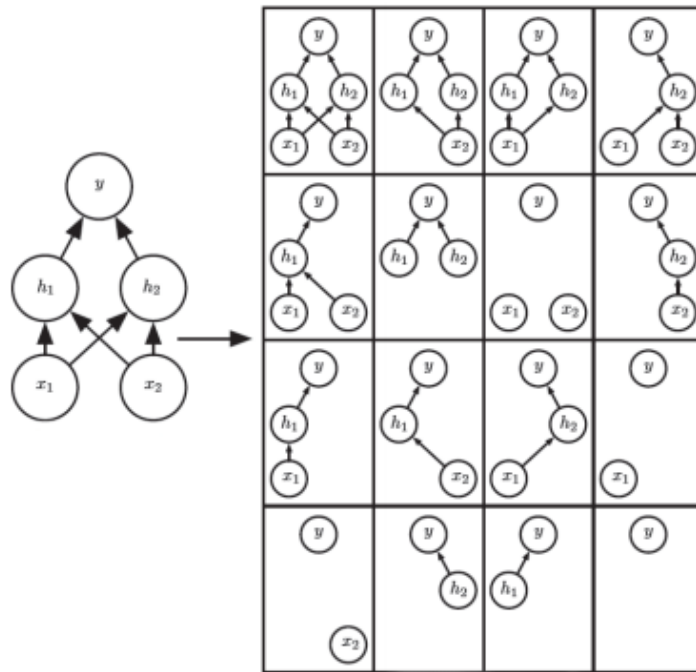
Еще один подход к созданию ансамблей - это обучение датасета на разных моделях. К таким подходам можно отнести:

- Dropout Ensemble
- Horizontal Voting Ensemble
- Snapshot Ensemble

# Dropout Ensemble

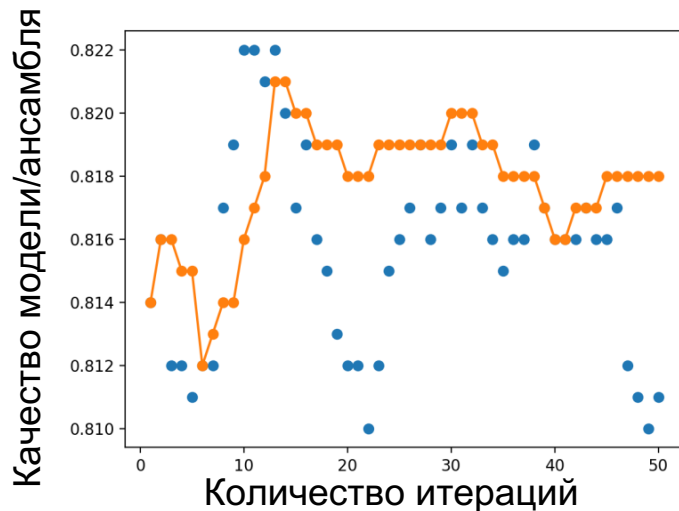
В процессе dropout обучается ансамбль, состоящий из подсетей, получаемых удалением невыходных блоков из базовой сети.

Невыгодно обучать на больших нейросетях.



# Horizontal Voting Ensemble

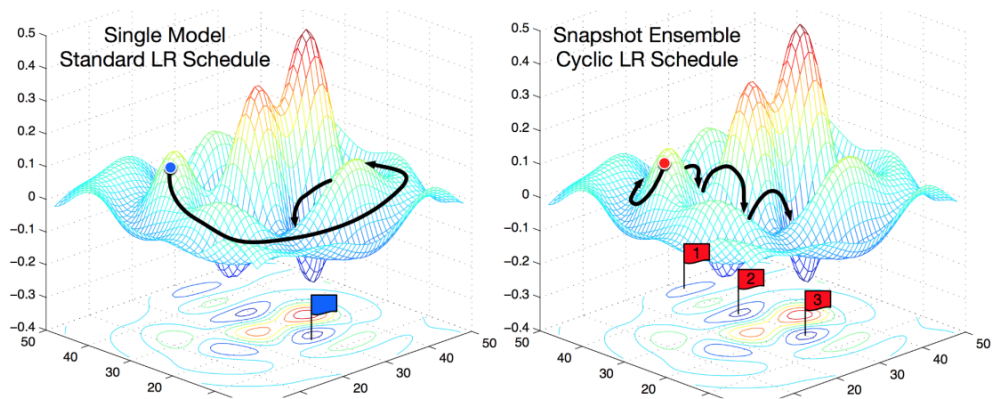
В этой модели ансамблей используются промежуточные результаты моделей не прошедших обучение до конца. То есть если в обучении модели 1000 итераций, можно, например, взять модели полученные на 900 и последующих 100 итерациях как новые части для ансамбля.



# Snapshot Ensemble

В Snapshot предлагается менять длину шага спуска в процессе обучения тем самым создавая расслоения моделей во время одного обучения.

SGD с агрессивным изменением шага спуска называется Stochastic Gradient Descent with Warm Restarts(SGDR).



# Компоновка результата

Остается последний элемент ансамбля нейронных сетей - вычисление финального результата из подсчитанных подсетями результатов. Вот некоторые способы получения конечного предсказания в ансамблях:

- Model Averaging Ensemble
- Model Weight Averaging Ensemble
- Stacking Ensemble
- Boosting

# Model Averaging Ensemble

Model Averaging Ensemble - самый простой из методов компоновки результатов в ансамбле. Берется среднее от всех предсказаний.

$$\hat{y} = \frac{\sum_{i=1}^n \hat{y}_i}{n}$$



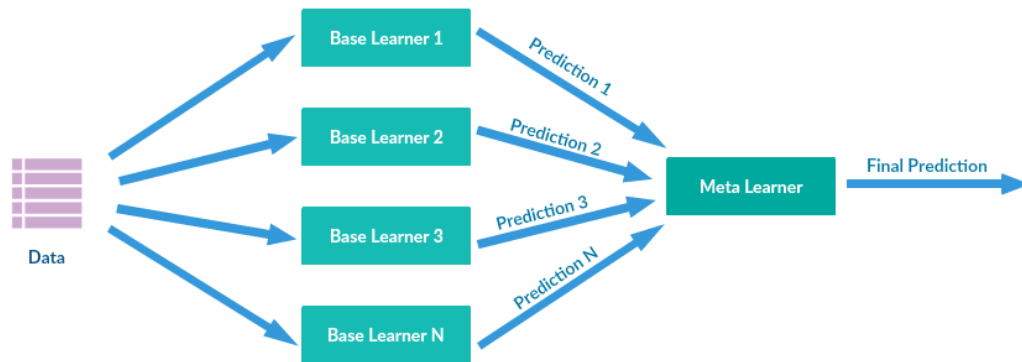
# Model Weight Averaging Ensemble

Идея этого метода в том что формируется новая нейросеть, веса в которой - среднее от весов по всем использованным в ансамбле моделям.

$$w_{ij} = \frac{\sum_{k=1}^n w_{kij}}{n}$$

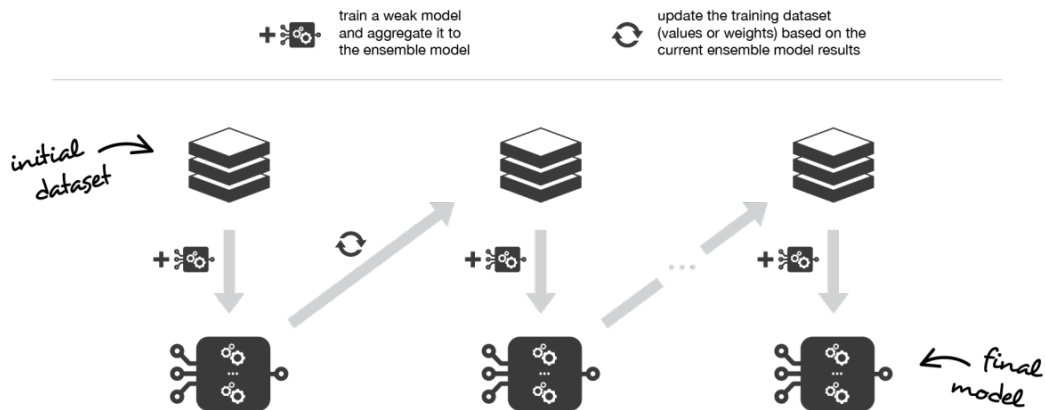
# Stacking Ensemble

Стэкинг заключается в добавлении к ансамблю мета-модели, которая получает на вход результат работы подсетей и обучается, таким образом присваивая каждой подсети свой вес. Финальный результат - линейная комбинация результатов работы подсетей.



# Boosting

Бустинг состоит в наслоении к модели небольших подмоделей учитывающих слабости нынешней модели, после чего модель обновляется.



# AdaBoost(Adaptive Boosting)

- В Adaboost ансамбль описывается как сумма из подмоделей с коэффициентами
- Обновление модели можно записать как
- Подбор коэффициента осуществляется путем решения задачи минимизации ошибки(E)

$$s_L(.) = \sum_{l=1}^L c_l \times w_l(.)$$

$$s_l(.) = s_{l-1}(.) + c_l \times w_l(.)$$

$$(c_l, w_l(.)) = \arg \min_{c, w(.)} E(s_{l-1}(.) + c \times w(.))$$

# AnyBoost(Gradient Boosting)

Как и в адаптивном алгоритме ансамбль представляет собой линейную комбинацию подмоделей с коэффициентами. Тут подмодели пытаются с использованием градиента уменьшить ошибку ансамбля.  $C$  - длина шага.

$$s_l(.) = s_{l-1}(.) - c_l \times \nabla_{s_{l-1}} E(s_{l-1})(.)$$

# Uncertainty Estimation

Очень важно избегать неправильных ответов, в которых нейросеть имеет большую уверенность, для этого имеет смысл знать степень уверенности модели в ответе.

Ответы ансамбля нейросетей используются для оценки уверенности модели в своем ответе.



# Вопросы

1. Зачем вообще понадобилось ансамблирование?
2. Что такое bagging и как он работает?
3. В чем заключается принцип работы Snapshot Ensembles?
4. Назовите два примера комбинации предсказаний в ансамбле.

# ИСТОЧНИКИ

- Про ансамбли верифицирующие данные - <https://machinelearningmastery.com/how-to-create-a-random-split-cross-validation-and-bagging-ensemble-for-deep-learning-in-keras/>
- Про ошибки - <https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/>
- Про ансамбли в целом + dropout- “Глубокое обучение” Я. Гудфеллоу, И. Бенджио, А. Курвилль
- Про Snapshot - <https://machinelearningmastery.com/snapshot-ensemble-deep-learning-neural-network/>
- Еще Snapshot - <https://medium.com/analytics-vidhya/snapshot-ensembles-leveraging-ensembling-in-neural-networks-a0d512cf2941>
- Еще Snapshot - <https://arxiv.org/pdf/1704.00109.pdf>
- Про HVE - <https://machinelearningmastery.com/horizontal-voting-ensemble/>
- Model averaging - <https://machinelearningmastery.com/model-averaging-ensemble-for-deep-learning-neural-networks/>
- Model Weight Averaging Ensemble - <https://machinelearningmastery.com/polyak-neural-network-model-weight-ensemble/>
- Stacking Ensemble- <https://machinelearningmastery.com/stacking-ensemble-for-deep-learning-neural-networks/>
- Boosting/Bagging - <https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205>
- Uncertainty Estimation - <https://arxiv.org/pdf/1907.06890.pdf>
- Uncertainty Estimation - <https://papers.nips.cc/paper/7219-simple-and-scalable-predictive-uncertainty-estimation-using-deep-ensembles.pdf>