Машинное обучение.
Домашнее задание №4

Задача 1. Убедитесь, что можете ответить на следующие вопросы:

1. Как измеряется качество метода обучения, если известно распределение $p(x, y)$?
2. Что такое $\mu(X^\ell)$ — число, функция, что-то еще? Что такое $\mathbb{E}_{X^\ell}[\mu(X^\ell)]$?
3. Какие формулы у шума, смещения, разброса? Какой смысл у этих компонент?
4. Приведите пример семейства с маленьким смещением и большим разбросом. Приведите пример семейства с большим смещением и маленьким разбросом.
5. Как сгенерировать подвыборку с помощью бутстрэпа?
6. Что такое бэггинг?
7. Как соотносятся смещение разброс композиции, построенной с помощью бэггинга, со смещением и разбросом одного базового алгоритма?
8. Как обучается случайный лес? В чем отличия от обычной процедуры построения решающих деревьев?
9. Почему хорошими базовыми алгоритмами для бэггинга являются именно деревья?
10. Как оценить качество случайного леса с помощью out-of-bag-процедуры?
11. Как случайный лес связан с метрическими алгоритмами? Как в нем измеряется сходство между объектами?

Задача 2. Пусть подвыборка $\tilde{X}^\ell$ генерируется с помощью бустрэппинга из выборки X^ℓ размера ℓ. Найдите вероятность того, что фиксированный объект $x \in X^\ell$ попадет в подвыборку $\tilde{X}^\ell$. Чему равна эта вероятность, если $\ell \to \infty$?

Задача 3. Известно, что бэггинг плохо работает, если в качестве базовых классификаторов взять методы ближайшего соседа. Попробуем понять причины на простом примере.

Пусть дана выборка X^ℓ из ℓ объектов с ответами из множества $\mathbb{Y} = \{-1, +1\}$. Будем рассматривать классификатор одного ближайшего соседа в качестве базового алгоритма. Построим с помощью бэггинга композицию длины N:

$$a_N(x) = \operatorname{sign} \sum_{n=1}^{N} b_n(x).$$

Оцените вероятность того, что ответ композиции на произвольном объекте x будет отличаться от ответа одного классификатора ближайшего соседа, обученного по всей выборке. Покажите, что эта вероятность стремится к нулю при $N \to \infty$.

Подсказка: ответ композиции на x может отличаться от ответа одного алгоритма только в том случае, если ближайший сосед x попал в обучение для менее чем половины базовых алгоритмов.

Задача 4. Пусть $x_1, \dots, x_N$ — одинаково распределенные случайные величины с дисперсией σ^2. Если они независимы, то дисперсия их среднего равна σ^2/N. Покажите, что если корреляция между любой парой этих величин равна $\rho > 0$, то дисперсия среднего вычисляется по формуле

$$\mathbb{D}\left[\frac{1}{N}\sum_{n=1}^{N} x_n\right] = \rho\sigma^2 + \frac{1-\rho}{N}\sigma^2.$$