

Время на выполнение: 20 минут

1. (2 балла) Количество мёда в горшках, которое Винни-Пух хочет съесть в гостях, можно представить в виде $y_i = 5 + u_i$, где u_i имеет равномерное распределение $u_i \sim \mathcal{U}[-2; 2]$, а i — номер визита в гости.

При прогнозировании аппетита Винни-Пуха Кролик абсолютно игнорирует все его прошлые визиты и просто подкидывает правильную монетку восемь раз. При каждом выпадении орла, Кролик ставит на стол очередной горшок мёда из глубоких запасов.

Постройте разложение квадратичной функции потерь прогноза Кролика на компоненты дисперсии, смещения и шума.

2. (1 балл) Герман делит дата сайнтистов на три типа: специалисты по deep learning (20%), machine learning (60%), big data (20%). Посчитайте индекс Джини и энтропию.
3. (1 балл) Постройте пример, в котором индекс Джини больше 0.99.
4. (2 балла) Постройте регрессионное дерево для прогнозирования y с помощью x на обучающей выборке. Узлы делятся до тех пор, пока в узле остается больше двух наблюдений.

| | | | | | | | |
|-----|-----|-----|-----|----|----|----|----|
| y | 100 | 102 | 103 | 50 | 55 | 61 | 70 |
| x | 1 | 2 | 3 | 4 | 5 | 6 | 7 |