Обобщим изученные выше схемы валидации в виде таблицы:

|  |  |  |  |
| --- | --- | --- | --- |
| **СХЕМА ВАЛИДАЦИИ** | **ПРЕИМУЩЕСТВА** | **НЕДОСТАТКИ** | **ФУНКЦИЯ В SKLEARN** |
| **hold-out** | Очень простой и понятный.  Данный метод чаще всего применяется в случае больших датасетов в силу того, что требует значительно меньше вычислительных мощностей, чем другие методы. | Разбиение производится случайным образом, и оценка в этом методе зависит от того, какие наблюдения попали в набор для валидации. | train\_test\_split() |
| **k-fold** | Более устойчивая к выбросам оценка качества модели.  Значения метрик получаются более объективными, чем в *hold-out*. | Обучаем одну и ту же модель  раз, что плохо сказывается на производительности. Если модель обучается медленно, то валидация может занять очень много времени. | KFold + cross\_validate() |
| **leave-one-out** | Идеально подходит для небольших датасетов (менее 100 примеров).  Поскольку все доступные данные используются как для обучения, так и для валидации, значения метрик наиболее объективны и надёжны. | Обучаем одну и ту же модель  раз.  Поэтому метод не подходит для оценки качества модели на больших наборах данных, так как становится очень ресурсозатратным. | LeaveOneOut + cross\_validate() |