Введение

Jacknife method подталкивает к обсуждению прикладных аспектов ресэмплинга и оценки погрешности. **Цель**: эмпирически оценить обобщающую способность алгоритма.Однозначный фаворит - Кросс-валидация, анализирующая эффективность построенной модель для выбора более оптимальных гиперпараметров из предложенных ('**pseudo-out-of-sample forecasts**'). Несмотря на многообразие техник CV, они все подпадают под общее описание:

$$\hat{\mathcal{L}}^{CV}(\mathcal{A}; D_n; (I_j^{(t)})_{1 \le j \le B}) = \frac{1}{B} \sum_{j=1}^{B} \hat{\mathcal{L}}^{HO}(\mathcal{A}; D_n; I_j^{(t)})$$

где $\mathcal{A}$ - статистический алгоритм, D_n - обучающая выборка, , $I_j^{(t)}$ - обучающее подмножество множества $\{1, ..., n\}$.

Типы CV

Базовый подход именуется **Hold-Out Validation**. Метод заключается в разделении данных на две выборки, поочередно становящиеся обучающей и валидационной. Среди преимуществ: отсутсвие затрат по времени и некоторый контроль переобучаемости. Однако столь наивное разбиение is highly risky.Возможна утеря информации об определенных классах, концетрация выбросов в одной выборке и так далее.
Поэтому самый известный и популярный алгоритм (imho) - это **k-Fold**. Предлагается разбиение обучающей выборки на k непересекающихся одинаковых по размеру частей. После происходит k итераций обучения, где каждый раз новая часть становится валидационной.Особого внимания достоен **stratified K-fold**, гарантирующий равные пропорции разбиения таргета в выборках.
Также существуют **exhaustive** методы: крайне объемные по вычислениям, но дающие большую точность. К примеру, **Leave-p-out CV**. Суть проста: выбираем размер валидационного подмножества $p \in [1, n]$. Далее - уже знакомый процесс при всех возможных разбиениях. Напомним, что всего получится C_n^p вариаций. И еще - что время деньги. С другой стороны, при $p = 2$ модель CV достигает наиболее непредвзятые результаты (при оценке **AUC-ROC**).

Риски CV

Здесь мы плавно переходим к ключевому недостатку CV - низкий уровень

надежности результатов при маленьких выборках (bias and variance).Приведем одну из техник контроля, подразумевающую использование некоторой предварительной оценки гиперпараметра λ_R. Концепция подразумевает обращение к следующей формуле, потенциально позволяющей противостоять волатильности показателей:

$$L_{\lambda_i} = \gamma \cdot \text{Relative Simplicity} + (1 - \gamma) \cdot \text{Relative Accuracy} \underset{\lambda}{\rightarrow} \min$$

где Relative Simplicity $= \frac{(\lambda_i - \lambda_R)^2}{(\lambda_{\max} - \lambda_R)^2}$ указывает на величину отклонения, а Relative Accuracy - на отношение определенных метрик качества для λ_i относительно pre-defined λ_R. Общая же идея: гиперпараметр γ определяет насколько велик должен быть прирост по эффективностич, чтобы допустить отклонение от некоторого базового значения λ_R. Таким образом, мы получаем возможность штрафовать сильные отклонения при маленьких выборках, меняя относительные веса прогнозов CV.