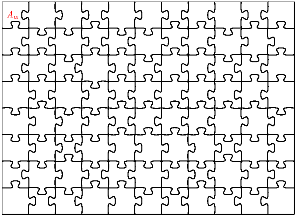
Statistical Machine Learning
Part 6

Horacio Gómez-Acevedo
Department of Biomedical Informatics

February 17, 2021



A_α

Resampling Methods for Parameter Estimation

Suppose you have applied your state of the art algorithm, but you don't know what is the distribution of a parameter (hyperparameter). The question is: how do I determine the bias and variance?

The **jackknife** and **bootstrap** are *resampling* methodologies that help improving classification.

Jacknife

It was introducied by Maurice Quenouille around 1950's. Let's start with an example as motivation for the use of the jacknife.
Example. Let's suppose that we have m independent random variables $X_1, \ldots, X_m$ that follow the same distribution. We can define the statistic $\overline{X}$ defined as $\frac{X_1+\cdots+X_m}{m}$. The question is what is the standard deviation of this statistic given a set of observed values $X_1 = x_1, \ldots, X_m = x_m$?
Following the definition of variance, we can determine

$$\widehat{\sigma}^2(\overline{X}) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (x_i - \overline{x})^2 \qquad (1)$$

That was simple enough, but what about calculating an estimate of the variance for other common statistics as *mode*, or *median* or other statistics?

Jackknife

Let's define the sample average of the data set deleting the jth variable as

$$\overline{X}_{(j)} = \frac{1}{m-1} \sum_{k \neq j} X_k$$

We also define the statistic that is the *average* of these averages

$$\overline{X}_{(\bullet)} = \frac{1}{m} \sum_{k=1}^{m} \overline{X}_{(k)}$$

The **Jackknife** estimate of the standard deviation is

$$\widehat{\sigma}^2_{Jack}(\overline{X}) = \frac{m-1}{m} \sum_{i=1}^{m} (\overline{X}_{(i)} - \overline{X}_{(\bullet)})^2 \tag{2}$$

It can be verified that (1) and (2) coincide; however, this process allows a generalization of this method.

Jackknife

One of the biggest advantages of the expression (2) is that when we have an estimator $\hat{\theta}(x_1, \ldots, x_m)$ of the statistic θ, we can actually estimate the variance of such estimator

$$\hat{\sigma}^2_{jack} = \frac{m-1}{m} \sum_{i=1}^{m} (\hat{\theta}_{(i)} - \hat{\theta}_{(\bullet)})^2,$$

where

$$\hat{\theta}_{(i)} = \hat{\theta}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m)$$
$$\hat{\theta}_{(\bullet)} = \frac{1}{m-1} \sum_{i=1}^{n} \hat{\theta}_{(i)}$$

Jackknife bias

It is also possible to obtain the **jackknife bias** estimation
Recall the definition of bias

$$bias = \theta - E(\hat{\theta})$$

The Jackknife estimate of bias is given by

$$bias_{jack} = (m-1)(\hat{\theta}_{(\bullet)} - \hat{\theta})$$

Bootstrap

In a common definition, a *bootstrap* data set is one created by randomly selecting m points (with replacement) from the training set $\mathcal{D}$.

For example if our training data set consists of the points $\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$, then a bootstrap could be

$$
\begin{aligned}
B_1 &= \{(x_1, y_1), (x_2, y_2), (x_3, y_3)\} \\
B_2 &= \{(x_1, y_1), (x_1, y_1), (x_2, y_2)\} \\
B_3 &= \{(x_2, y_2), (x_3, y_3), (x_2, y_2)\}
\end{aligned}
$$

Bootstrap

In the bootstrap setup, the data sets (say B_js in our example) are treated as independent sets. The **bootstrap** estimate of a statistic θ is defined as

$$\hat{\theta}^{*(\bullet)} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^{*(b)},$$

where $\hat{\theta}^{*(b)}$ is the estimate of θ for the sample b .

Bootstrap bias and variance estimates

The bootstrap estimate of the bias

$$bias_{boot} = \hat{\theta}^{*(\bullet)} - \hat{\theta}$$

Whereas the bootstrap estimate of the variance is

$$\hat{\sigma}^2(\theta) = \frac{1}{B} \sum_{b=1}^{B} \left(\hat{\theta}^{*(b)} - \hat{\theta}^{*(\bullet)} \right)$$

References

Materials and some of the pictures are from (1),(2), and (3).

1. Gareth James et al. *An Introduction to Statistical Learning with applications in R*. Springer (2015)
2. Richard O. Duda et al. *Pattern Classification* John Wiley (2001).
3. Aurélien Géron. *Hands-on Machine Learning with Scikit-Learn & TensorFlow* O'Relly (2017)
4. Wiebe R. Pestman *Mathematical Statistics* de Gruyter (1998)
5. Bradley Efron. *The Jackknife, the Bootstrap and other Resampling Plans* SIAM (1982)

I have used some of the graphs by hacking TiKz code from StakExchange, Inkscape for more aesthetic plots and other old tricks of TEX