Marco Cerliani · Follow

Feb 24, 2021 · 6 min read · ▶ Listen

Advanced Permutation Importance to Explain Predictions

Bring Explainability to the Next Level preserving Simplicity

Photo by Xavi Cabrera on Unsplash

Machine Learning interpretability is an active field of research that involves all the techniques useful to provide more informative predictions. Predictive models have fame to be considered as black-box instruments optimized only to maximize the performances. Accuracy is important but, in most business cases, inspect why a

much some features affect the predicted values. Practically we are doing inference on our model, following some heuristics, and then we try to put all the information inside a table or better into an awesome graph.

When we are trying to explain the output of any machine learning model, we should take into consideration a crucial aspect. **Interpretability may not result in explainability.** The adoption of the latest techniques in the field or the production of the coolest graphs may be useless if the people haven't adequate knowledge to understand it. This is the case if we have to show the outcomes to the business unit. An overloaded graph or the usage of some complex indexes, which aren't understandable to all, make our work not explainable.

In this post, we provide some graphical reports to explain the output of machine learning models. We leverage the simplicity and the adaptability of the permutation importance to provide different graphical reports, which can also be compared to the outcomes obtained with the SHAP approach.

PERMUTATION IMPORTANCE: BASIC USAGE

Permutation importance is a frequently used procedure for feature importance computation that every data scientist must know. This technique aims to alter the relationship between the target and the features. The permutation implies the drop in the model performances which is indicative of how much the model depends on the feature.

In detail, the permutation importance is calculated as follows. First, a model is fitted and a baseline metric is calculated on some data. Next, a feature from the same data is permuted and the metric is evaluated again. **The permutation importance is defined to be the difference between the permutation metric and the baseline metric**. These steps are computed for all the columns in the dataset to obtain the importance of all the features. A high value means that the feature is important for the model. In this case, the shuffling of the values brokes the relationship with the target and results in low-quality predictions (high error). Instead, a low value means the permutation metric is near to the original one, i.e., a
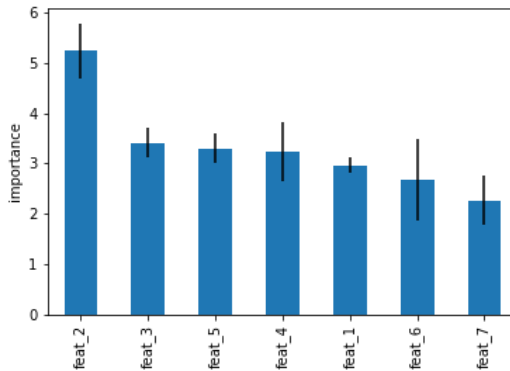
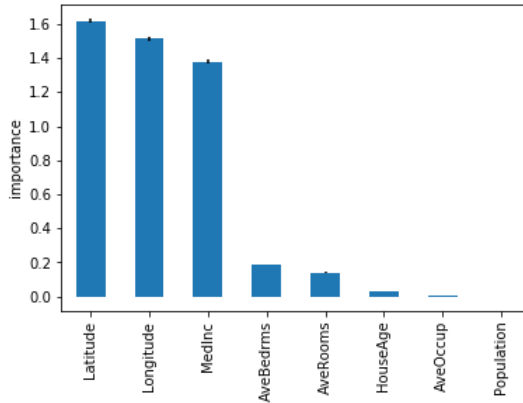Example of permutation importance for feature evaluation (image by the author)

The graph above is very common when computing permutation importance. We report, in descending order, the mean permutation score (obtained with several permutations runs) together with the relative standard deviations in the form of error bars. Most of the time, the usage of permutation importance ends here, showing the previous bar plot. This graph is awesome, it says us a lot of information and it's easily understandable by all. Can we make some steps further? Can we take advantage of the simplicity of permutation importance to provide more detailed explanations of our predictions?

What we try to do now is to provide a **permutation importance score** for the observations of our interest. In this way, we have a more detailed view of the decisions made by our model. We leverage the simplicity of permutation importance to show how features contribute to the prediction of each sample. This analysis is easily accessible for every kind of supervised task (regression or classification) and leads to produce some awesome graphs.

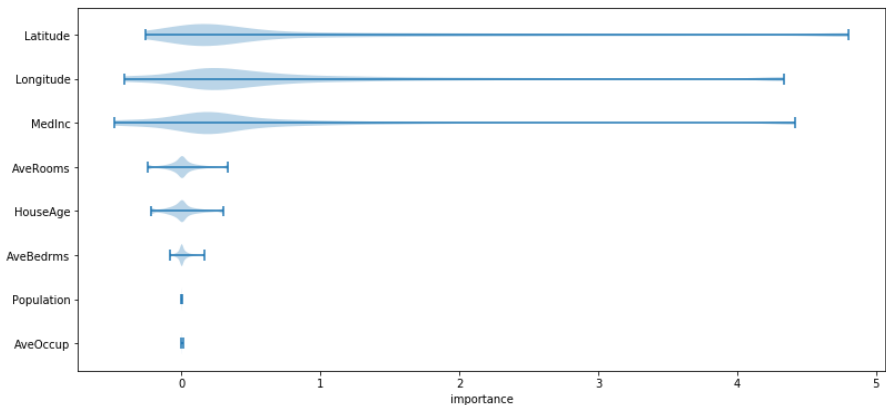## PERMUTATION IMPORTANCE: ADVANCED USAGE

Let's start in a regression scenario. We are carrying out a regression task and we are interested in predicting the house values given some external numerical features. After fitting a model of our choice we easily compute the permutation importance
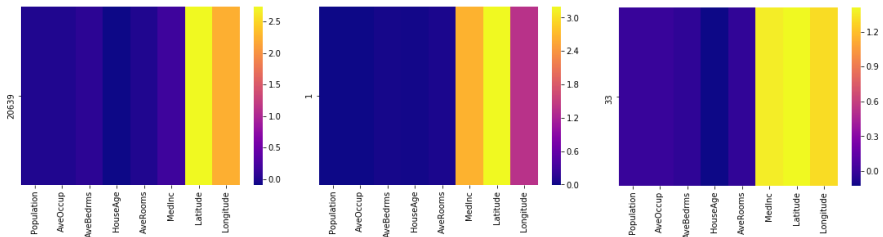
Permutation importance for feature evaluation in a regression task (image by the author)

In this case, the importance is computed as deviations from the mean squared error. We repeat the same approach but without averaging the score for columns. We simply compare the squared errors of the original predictions with the ones obtained permuting the features per sample. In this case, we summarize the sample scores, obtained from multiple repetitions, taking the median value. Following the procedure, we end with an importance score for each observation in each column. We observe the distributions of the sample scores below.

heatmaps below, we show how much each feature affects the predicted values for some random instances.

Permutation importance for individual sample (image by the author)

To get a more general view, we can plot the effects on the whole data for the desired features. For example, low values of Latitude and high values of Longitude have a high impact on the predicted values.
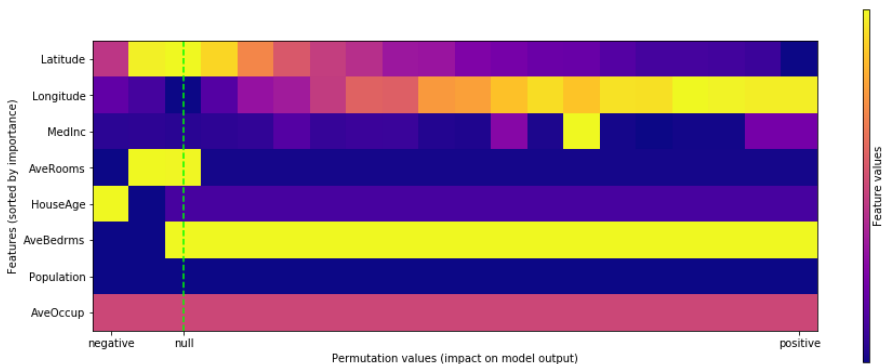
Summarize the effects of all the features (image by the author)

The same representation can be computed in 2D to visualize also interactions between features or between features and target.
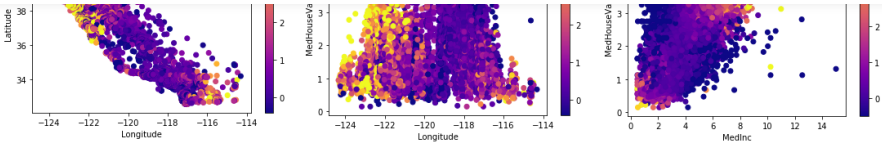
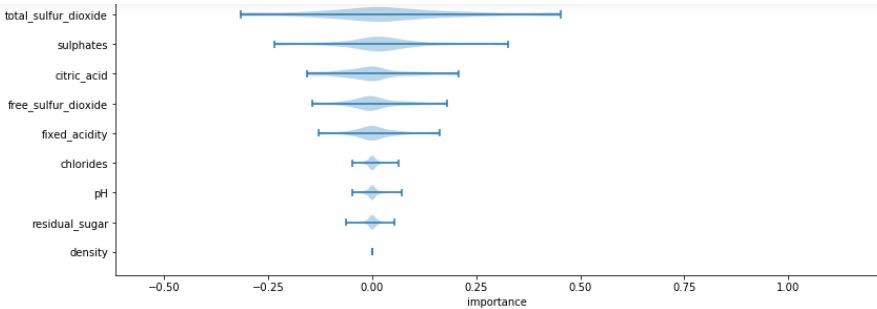Permutation importance for samples and multiple features (image by the author)

For classification tasks, we operate the same reasoning and graphical representations. Let's say we are interested in predicting the quality of wines given some external numerical features. After fitting a model of our choice we easily compute the permutation importance. We can't use the mean squared error as a scoring function for the permutation importance. A valuable alternative, in this case, is the log loss (categorical cross-entropy to handle multiple classes). What was introduced before remains valuable also in this scenario.
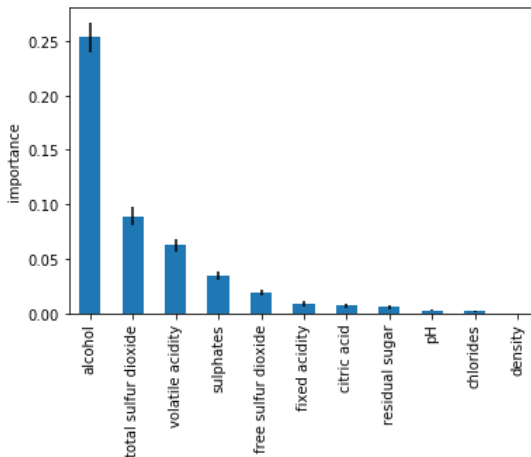
Permutation importance for feature evaluation in a classification task (image by the author)
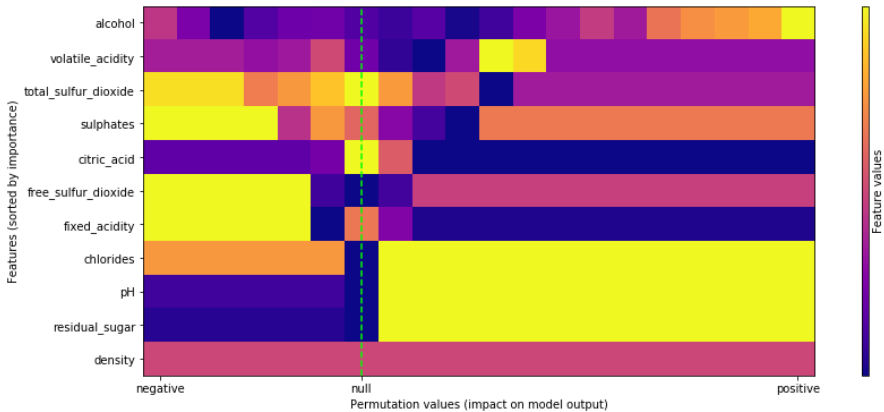
We retrieve as before the importance score per sample. Here the negative effect of some observations is more evident than in the previous case.

Permutation importance for sample evaluation in a classification task (image by the author)
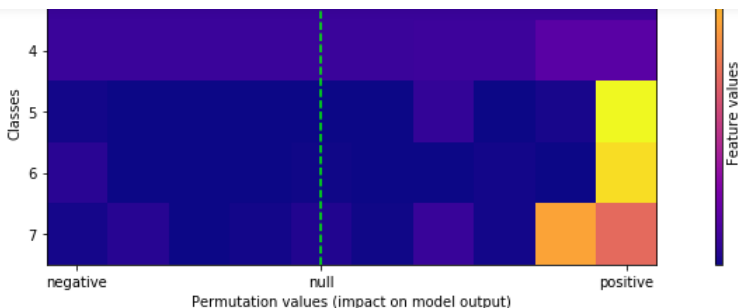
Again, it's possible to show the importance, in heatmap format, for each sample and all the samples in each feature.

Summarize the effects of all the features (image by the author)

Maintaining the same visualization format, it's possible to show the impact of the features on the predicted classes.

Summarize the effects of a feature on all the classes (image by the author)

For example, low values of volatile acidity have a great impact on labels 3 and 4, while for labels 7, 6, and 5 high values of acidity have a great impact.

SUMMARY

In this post, we introduced the basic concepts of permutation importance as a procedure for feature importance computation. Then we tried to leverage its simplicity adapting it to provide better explanations of our model outputs. As result we obtain useful insights to show in a graphical format, preserving the explainability also for not technical people.

CHECK MY GITHUB REPO

Keep in touch: Linkedin

Sign up for The Variable

Get this newsletter

Emails will be sent to jeabinay@gmail.com.
Not you?