Support vectors

# Basics:

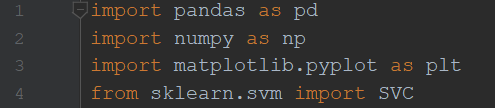


Figure 1. Imports

Again, we will store dataframe from **titanic.csv** in variable named **titanic**, **feature matrix** in **x** and **target array** in **y**, **support vectors classification model** in **model**

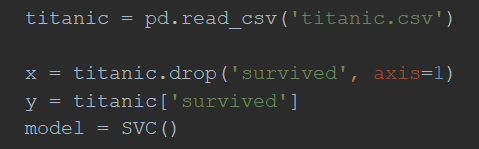


Figure 2. Main variables

This time, before estimating our model, we need to set some parameters. We will experiment with *‘C’* and *‘gamma’*. To do this we import **validation\_curve** (visualization the relation between model’s accuracy and parameter value) and **GridSearchCV** (looks for the best parameters values within the given range).



Figure 3. Tools for parameters setting

# *‘C’* and *‘gamma’* setting:

1. Let’s start from *‘gamma’*. It may be a **float** value, **‘auto’** or **‘scale’**. Firstly, we compare **‘auto’** and **‘scale’** with **C=1.**0 (default value) using **valitadion\_curve** (I decided to visualize it as bar, because we have only two values). Secondly, we compare float numbers between **0.001** and **0.1** (for creating this numbers we will use **linspace()** from **numpy**) with **C=1.0**. This is not the best way to configure parameters, because we do our research with fixed *‘C’* value. Soon we will see another way of setting the parameters, which is very easy to use and, for small ranges of changes, fast enough.

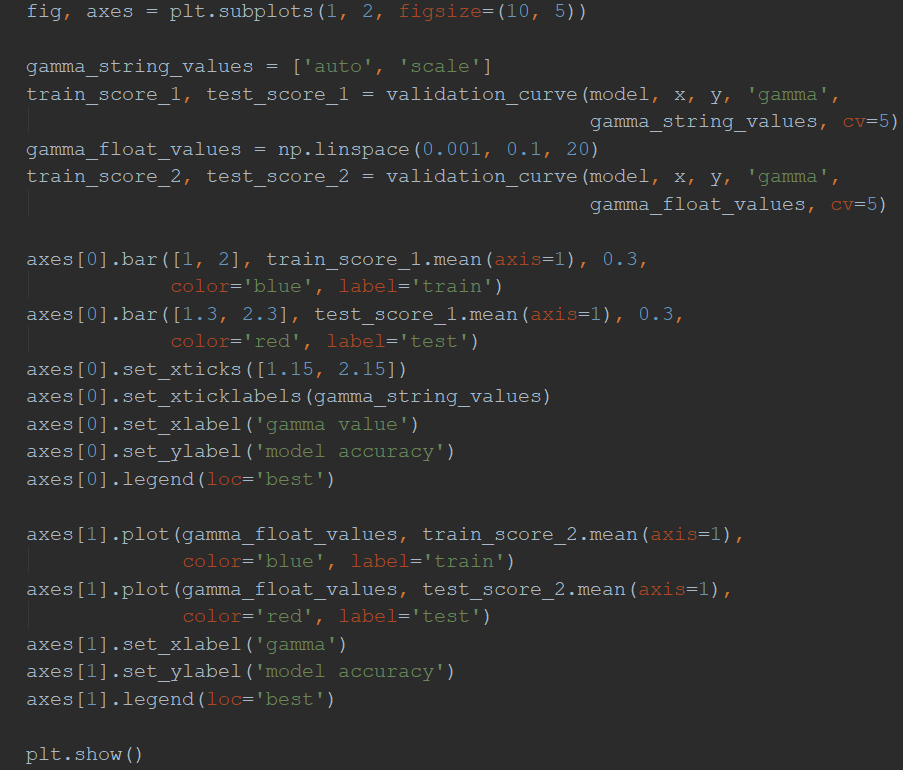
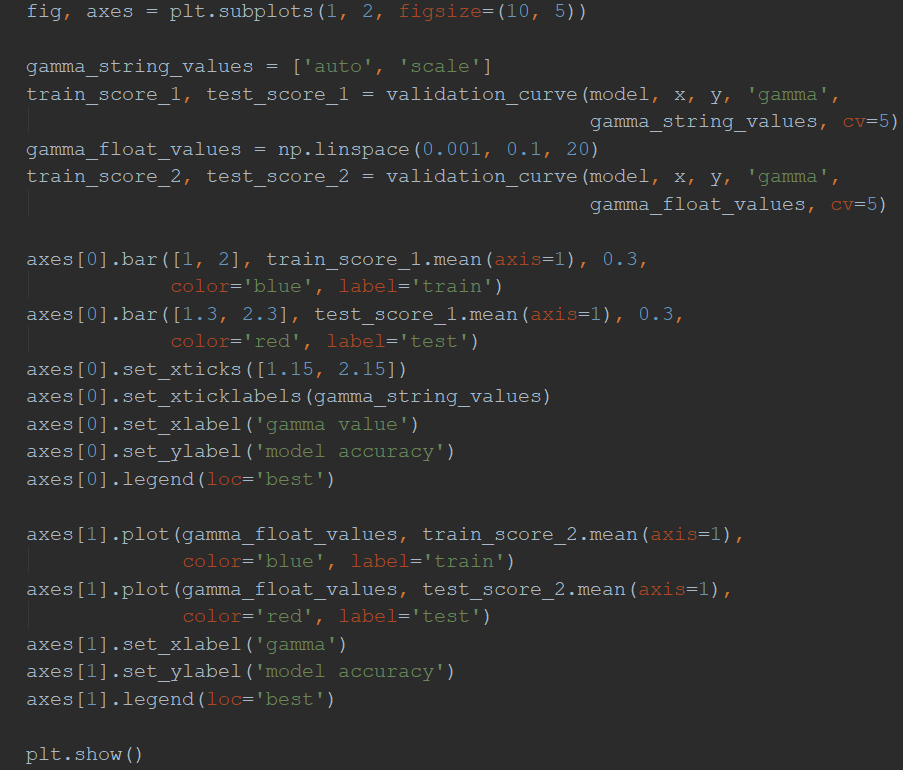


Figure 4. Plotting graphs

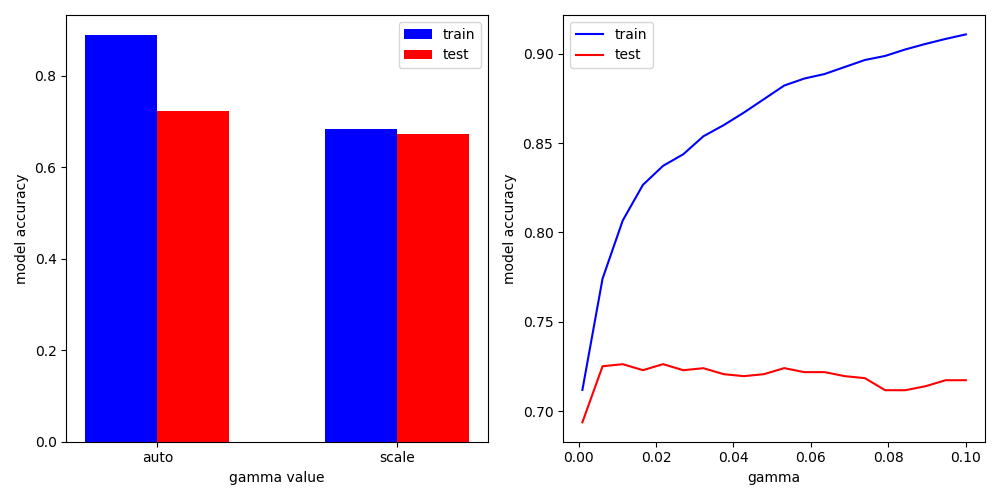


Figure 5. Gamma/accuracy relation

The result isn’t as good as it was with **GaussianNB** model. But here we configured only one parameter. In the next point, we will set **both** of them and calculate the accuracy of our current model.

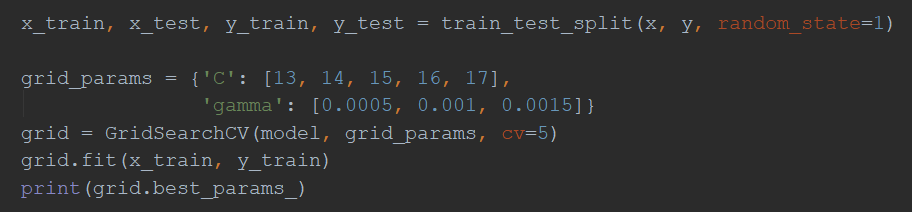
1. For setting *‘C’* and *‘gamma’* parameters together we will use **GridSearchCV**.

Figure 6. Using grid\_search

After printing **grid.best\_params\_** we will see **C = 15**, **gamma = 0.001**. These values are in the middle of our ranges, so we don’t need to change them (if the best values are on the borders, it means that truly best values can be out of out range). Now we can test our model on **x\_test** and **y\_test** the same way we did with **GaussianNB**:

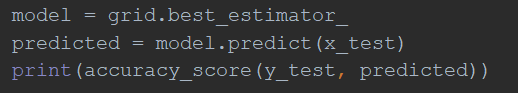


Figure 7. Predicting with x\_test

The accuracy we get there is **0.78**. It isn’t better than the result we got before. So, **FOR THIS ‘TITANIC’ DATASET**, from **GaussianNB** and **svc** models it is more meaningful to choose the first one, because **naive\_bayes** models are easier to set and use.