

[CM2] To begin understanding the dataset, generate a \"pairs plot\" (also called a scatter plot matrix, seaborn.pairplot is one method to do this) of the data. Note that the pairs plot includes the scatter plots of every dimension versus every other dimension. From the pair plot, identify the subplots corresponding to the pairs of features where you see correlation.

IRIS DATASET

For Iris: Make a single pair plot of all the features and data

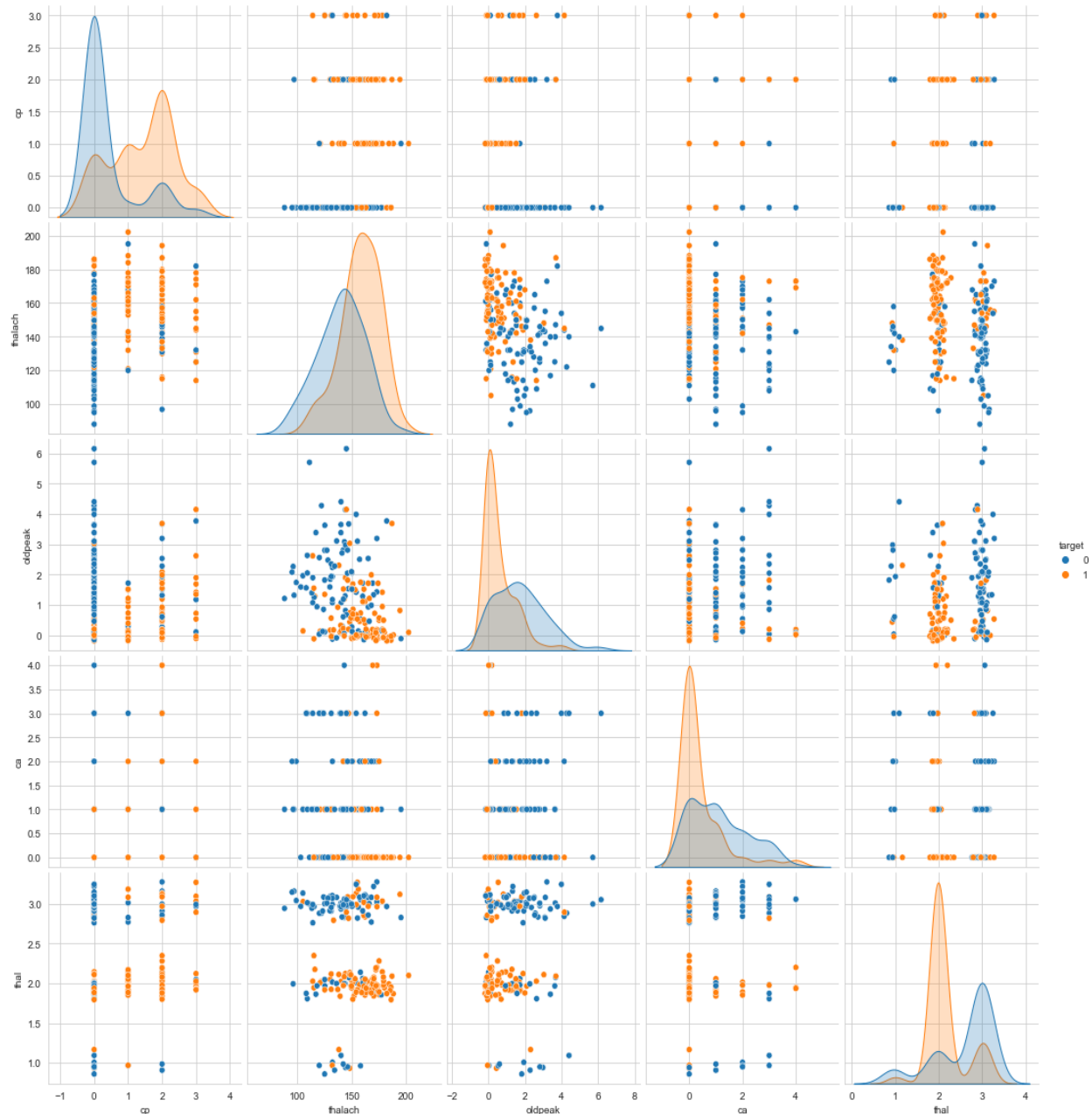
```
In [12]: plt.close();
sns.set_style(\"whitegrid\");
sns.pairplot(df_irisnew, hue='species', height=3);
plt.show()
```



- In petal_length Vs sepal_length and petal_width Vs sepal_length we can separate setosa flowers from non-setosa flowers very easily by creating decision boundary.
- From scatterplot of petal_length Vs petal_width we can clearly see the separation of different species without much overlap so these are the most useful features to classify various flower types.
- Here setosa can be easily identified (linearly separable) versicolor and virginica has some overlap (almost linearly separable).
- To classify the flower types we can find lines and if-else condition to build a simple model.

HEART DISEASE DATASET

For Heart Disease: Use pair plots to explore different subsets of features and their correlations, distribution, etc. You don't need to include this large pair plot in your report. Instead, choose your own subset of 3-5 features for the plot which highlight some interesting pattern. Show this smaller pair plot only and provide a short justification why you chose those features



- From the pair plot given above we can clearly see that no pair of features can clearly distinguish between the targets.
- Using thal vs thalach and thal Vs oldpeak we can observe clustered pattern of targets.
- However ca Vs cp can separate the target up to reasonable extent