# ML metrics (good for understanding “Classification report” from SciKit Learn

<https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_algorithms_performance_metrics.htm>

# Calendar Description automatically generated

# Multiclass vs MultiLabel problems

<https://stats.stackexchange.com/questions/11859/what-is-the-difference-between-multiclass-and-multilabel-problem>

Multiclass: classes are mutually exclusive  
Multilabel: classes may have some cross over relations

“In the famous leptograspus crabs [dataset](http://www.stats.ox.ac.uk/pub/PRNN/) there are examples of **males and females** of **two-colour forms** of crab.  
You could approach this as a :

**multi-class problem** with **four classes** (male-blue, female-blue, male-orange, female-orange)

or

as a **multi-label problem**, where **one label would be male/female** and the **other blue/orange**.   
Essentially in multi-label problems a pattern can belong to more than one class.”

Our Exoskelly data problem is multiclass.

# SVM - Decision\_function\_shape:

* ovr = one vs Rest
* ovo = one vs one

<https://stackoverflow.com/questions/43505451/which-decision-function-shape-for-sklearn-svm-svc-when-using-onevsrestclassifier>

# ROC\_AUC for Multiclass problem:

ROC\_AUC: need to plot as “one vs rest” rather than “one vs one”?

<https://stats.stackexchange.com/questions/2151/how-to-plot-roc-curves-in-multiclass-classification>

<https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html>

# What is ROC AUC?

## First ROC

ROC is a plot between True Positive vs False Positive Rates. This makes the metric fitted to binary problems.

Either the event is present or not and you got the classification right or wrong.

Table

Description automatically generated

**False positive** a test result which wrongly indicates that a particular condition or attribute is present.

Chart, line chart

Description automatically generated

The centre blue line shows a purely random guess classifier (50% hit or miss)

The more the curve deviates from the line, the better your performance.

## Thus AUC

Given the measurement of each axis is 1.0. Finding the area under the curve allows for a measure of how well your model performs.

Where 0.5 is really bad, 🡪 1 is amazing and   
🡪 0 is debatably also “amazing” (detecting precisely the opposite)

# ROC\_AUC score function

<https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html>

Graphical user interface, text, application, email

Description automatically generatedText

Description automatically generated Text

Description automatically generated

A picture containing text, computer, file

Description automatically generated

<https://discuss.analyticsvidhya.com/t/what-is-the-difference-between-predict-and-predict-proba/67376>