School of Computing and Information Systems
The University of Melbourne
COMP30027 MACHINE LEARNING (Semester 1, 2019)

Practical exercises: Week 5

Today, we will expect you to be referring to the API[1] for `scikit-learn` (`http://scikit-learn.org/stable/modules/classes.html`) — you should also refer to previous weeks' exercises where necessary. In particular, we will be examining the behaviour of some Support Vector Machine classifiers.

1. The `scikit-learn` documentation isn't just an API; there are many useful examples in the "User Guide" — navigate to "Plot different classifiers in the iris dataset" in Section 1.4.1. (`http://scikit-learn.org/stable/auto_examples/svm/plot_iris.html`).

   (a) Examine the visualisations of the four different[2] SVMs, paying close attention to the **decision boundaries**. Which do you think has the best expressitivity, based on the two–dimensional slice shown?

   (b) (Assuming that you can display the plots[3]): examine the visualisation of the four models on some other pairs of attributes from the *Iris* dataset, e.g. petal length vs. petal width.

   (c) The default value of the `C` parameter (the "penalty" for mis-classified examples) is 1. Increase (or decrease) this value and observe how the decision boundaries change.

   (d) Calculate the **training accuracy** of the various SVM classifiers that you graphed above. Do you see any evidence that any of these classifiers might be **overfitting** this data?

2. Let's summarise some earlier work. Use all four attributes from the *Iris* data[4], and compare the training accuracy with the accuracy estimated by **10–fold (stratified) cross–validation**, for the following models:

   (a) One-R

   (b) 1-Nearest Neighbour (`neighbors.KNeighborsClassifier`)

   (c) 5-Nearest Neighbour

   (d) Decision Trees[5]

   (e) `LinearSVC()`

   (f) SVMs with a cubic kernel

   (g) SVMs with an RBF kernel

[1] Note that there are probably some small differences between the version installed in the labs and the latest stable version.
[2] Well, the default implementation of `LinearSVC()` and `SVC(kernel='linear')` are slightly different, even if the basic method is the same.
[3] You might need to change `plt.show()` to `plt.savefig('/path/to/output.png')` to examine them.
[4] I know that *Iris* is pretty boring, but we'll look at some non-trivial datasets next time.
[5] `scikit-learn` uses CART instead of ID3/C4.5, but don't worry too much about this.