

## Homework 5.

### Fall 2019.

A) 80 points.

The jupyter lab file "HW5.ipynb" contains the knn multi-class classification example and B-cell tumour data from the last lecture on predictive models.

It contains a function "cross\_val\_score\_solution()" which computes a cross-validation estimate of accuracy. This function simply calls the corresponding function in the scikit-learn library.

In this programming exercise, you should rewrite this function using only functions and data structures in numpy and those built into python (and not using any from scikit-learn).

Your function should be fully commented and marks will be given for this.

(Also included are some example functions from numpy that you might find useful).

Upload your completed jupyter lab file as HW5\_NETID.ipynb and the corresponding html output file.

B) 20 points.

Now rewrite the above jupyterlab file as a .py file that can be run on the unix command-line. It should take a single command line argument that indicates how many folds to use in the cross validation.

Run the above script on the prince cluster, using job arrays to run the script in parallel on the prince cluster for 10, 5 and 3 fold cross-validations.

Upload your shell script used for this as HW5\_NETID.sh; your python script as HW5\_NETID.py; and your final output as HW5\_NETID.txt.