LING572 Hw8: SVM
Author: Dhwani Serai

Q1 (15 points): Run libSVM on a **binary** classification task.

Table 1: Results on the binary task

Expt id	Kernel	gamma	coef0	degree	total_sv	Training Acc	Test Acc	Test Acc from Q2
1	linear	-	-	-	535	99.722%	95%	95%
2	polynomial	1	0	2	792	99.722%	92%	92%
3	polynomial	0.1	0.5	2	775	99.722%	96.5%	96.5%
4	RBF	0.5	-	-	1798	99.722%	50%	50%
5	sigmoid	0.5	-0.2	-	1214	53.667%	40.5%	40.5%

Q2 (60 points): Write an SVM decoder, **svm_classify.sh**, that uses a SVM model created by libSVM to classify test instances.

- The command line is: svm_classify.sh test_data model_file sys_output

- The classifier should be able to handle the four types of kernels specified in Table 1. That is, it should be able to read the kernel type and parameters from the model_file and calculate the kernel function accordingly.

- test_data is in the libSVM data format (e.g., **test**).

- model_file is in the libSVM model format (e.g., **model_ex**). The model file stores $\alpha_i y_i$ for each support vector and ρ (See Slides #11-13 in class15_libSVM.pdf).

- Each line in sys_output (e.g., **sys_ex**) has the format "trueLabel sysLabel fx": trueLabel is the label in the gold standard, sysLabel is the label produced by the SVM classifier, fx is the value of $f(x) = wx - \rho = \sum_i \alpha_i y_i K(x_i, x) - \rho$.

  If $f(x) >= 0$, then sysLabel should be **0**; else sysLabel should be **1**. This is different from the convention used in SVM papers/chapters. For other differences between the two conventions, see slide #11 in class15_libSVM.pdf.

- Use the model file created in Q1 and **test** as the test data. Fill out the last column of Table 1. Save the sys_output file as sys.id, where id is the expt id in the first column of Table 1.

- You only need to submit **sys.1** and **sys.4**.