LING572 Hw8: SVM
Due: 11pm on March 1, 2022

The example files are under dropbox/21-22/572/hw8/examples/.

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

(a): The data are under **hw8/examples/**:

- **train.txt** and **test.txt** are the training and test data in the Mallet format.
- **train** and **test** are the data in the libSVM format.
- You only need to use **train** and **test** for this hw.

(b): Run libSVM with **train** as training data, **test** as test data, and the settings specified in the 2nd-5th columns of Table 1. Fill out the 6th-8th columns of Table 1. Save the model under model.id, where id is the expt id, specified in the first column.

(c): You only need to submit **model.1** and **model.4**.

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	-	-	-				
2	polynomial	1	0	2				
3	polynomial	0.1	0.5	2				
4	RBF	0.5	-	-				
5	sigmoid	0.5	-0.2	-				

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 slide #12-14 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 #14 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**.

Submission: Submit the following to Canvas:

- Your note file *readme.(txt | pdf)* that includes Table 1, and any notes that you want the TA to read.
- hw.tar.gz that includes all the files specified in dropbox/21-22/572/hw8/submit-file-list, plus any source code (and binary code) used by the shell scripts.
- Make sure that you run **check_hw8.sh** before submitting your hw.tar.gz.