Naive Bayes and Classifier Evaluation

1 Naive Bayes

You are building a naive bayes model for language detection using character bigrams, with the following training data:

Language	Sentence	Character Bigrams
english	learn	le ea ar rn
english	ablaze	ab bl la az ze
spanish	hablo	ha ab bl lo

1.1 N-Gram Counts (raw)

Record the counts of each character n-gram for each language below:

	ab	ar	az	bl	ea	ha	la	le	lo	rn	ze
spanish											
english											

1.2 N-Gram Counts (smoothed)

Now in the table below, convert your raw n-gram counts to smoothed versions, using Laplace (add-one) smoothing.

	ab	ar	az	bl	ea	ha	la	le	lo	rn	ze
spanish											
english											

1.3 Classify!

You want to classify a new single-word sentence, s: `able`.
Calculate $P(spanish|s)$ and $P(english|s)$. You should ignore unknown character bigrams, and use add-one smoothing.

$P(spanish|s) =$

$P(english|s) =$

2 Classifier Evaluation

You built a sentiment classification model, and get the following results on your test set:

Actual Label	Predicted Label
pos	pos
pos	pos
neg	pos
pos	pos
pos	pos
pos	pos
pos	pos
pos	pos

2.1 What is the accuracy of your model?

2.2 Is this a *good* model? Why or why not?

2.3 Fill in the confusion matrix for your model:

		Actual	
		pos	neg
Predicted	pos		
	neg		

2.4 Compute additional statistics based on the confusion matrix:

True Positives (tp)	False Positives (fp)	False Negatives (fn)	Precision (p)	Recall (r)